IJCSIS Vol. 8 No. 4, July 2010
ISSN 1947-5500

International Journal of Computer Science & Information Security

Editorial
Message from Managing Editor

The International Journal of Computer Science and Information Security is a monthly periodical on research articles in general computer science and information security which provides a distinctive technical perspective on novel technical research work, whether theoretical, applicable, or related to implementation.

Target Audience: IT academics, university IT faculties; and business people concerned with computer science and security; industry IT departments; government departments; the financial industry; the mobile industry and the computing industry.

Coverage includes: security infrastructures, network security: Internet security, content protection, cryptography, steganography and formal methods in information security; multimedia systems, software, information systems, intelligent systems, web services, data mining, wireless communication, networking and technologies, innovation technology and management.

Thanks for your contributions in July 2010 issue and we are grateful to the reviewers for providing valuable comments. IJCSIS July 2010 Issue (Vol. 8, No. 4) has an acceptance rate of 36 %.

Available at http://sites.google.com/site/ijcsis/

IJCSIS Vol. 8, No. 4, July 2010 Edition

ISSN 1947-5500 © IJCSIS 2010, USA.

Abstracts Indexed by (among others):

IJCSIS EDITORIAL BOARD

TABLE OF CONTENTS

Expert-Aware Approach: An Innovative Approach To Improve Network Data Visualization

Doris Hooi-Ten Wong
National Advanced IPv6 Centre (NAv6)
Universiti Sains Malaysia
11800, Penang, MALAYSIA
doris@nav6.org

Kok-Soon Chai
National Advanced IPv6 Centre (NAv6)
Universiti Sains Malaysia
11800, Penang, MALAYSIA
kschai@nav6.org

Sureswaran Ramadass
National Advanced IPv6 Centre (NAv6)
Universiti Sains Malaysia
11800, Penang, MALAYSIA
sures@nav6.org

Nicolas Vavasseur
Université de Franche Comté
16 route de Gray
25030 Besançon cedex, FRANCE
nicolas.vavasseur@edu.univ-fcomte.fr

***Abstract*—Computers have been infected by the computer anomalies. The availability of network data visualization tools greatly facilitate to perceive computer users from being affected by these anomalies. Many of the network data visualization tools are designed particularly for users with advanced network knowledge even though the tools are indispensable by diverse computer users. We proposed an expert-aware approach to designing a system which formulated with a large amount of network data and adaptive for diverse computer users. In the preliminary phase, we construct an intelligent expertise classification algorithm which provides a default setting for the expert-aware network data visualization tool. Besides, the tool will learn from continual user feedbacks in order to statistically satisfy the needs of majority tool users. In this paper, we will focus on the expert-aware approach with the users' expertise level in network security and adapts the visualization views that are best suitable for the computer user. Our initial results from the approach implementation showed that it is capable of representing several of network security data not only from small network but also for complicated high dimensional network data. Our main focus in this paper is to fulfill different requirements from diverse computer users.**

Keywords- network data visualization tool, network knowledge, expert-aware approach, network security.

I. INTRODUCTION

The evolution of hardware technology resulted in ton of data being captured and stored. Large volume of network data is being requested by diverse computer users. The network data are represented to computer users by using different kinds of existing network data visualization tools. Nowadays, many computers have been infected with the computer anomalies. The availability of network data visualization tools greatly facilitated to detect, perceive and defend computer users from being affected by these anomalies. This definitely entailed enormous network data visualization tools to completely represent network security data to the computer users. However, many of the network data visualization tools are designed particularly for users with advanced network awareness, although the tools are indispensable by various types of computer users. There are numbers of network data visualization tools that perform network security data in their respective way such as, bar graph, pie chart and others data visualization techniques. The network data are easily represented to users by using a bar chart or pie chart if they are a small amount, but very difficult for beginner computer user to understand the data structures information [1]. An intelligence approach shall come into the priority in order to improve the network data visualization. A scalable and intelligence expert-aware approach works by representing the network data in a more comprehensive way, effectively combining maximizing level of understanding among diverse computer users.

In Section II of this paper, we presented existing network data visualization tools and problems. In Section III, we discussed the architecture of the expert-aware approach. Finally, we discussed comparisons between expert-aware approach and existing approaches in section IV. The expected results of the proposed method and the contributions will be made in Section V and following by a conclusion of the paper in Section VI.

II. EXISTING NETWORK DATA VISUALIZATION TOOLS AND PROBLEMS

There are number of tools in the visualization area that have applied on the network data visualization. Commonly, network security data monitoring is the part that most of the visualization applications have been focused on more compared with others. Information on malicious attacks that have been triggered by using an abnormal detection device will be presented to the network administrators [2]. There are some other areas that visualization tools have focused on such as network intrusion detection and general network traffic. In this section, we discussed eight existing network data visualization tools which consist of network data and network security visualization tool. Network data visualization tools namely, WatchPoint, ntop, Nodemap while network security visualization tools are VISUAL, SCPD, PortVis, NVisionIP and NIVA.

A shorter version of this paper will appear in Proceedings of 2nd International Conference on Network Applications, Protocols and Services (published by IEEE Conference Publication Services), 22-23 September 2010, Alor Setar, Kedah Darul Aman, MALAYSIA.

A. Network Data Visualization Tools

1) Watch Point: WatchPoint is designed for presenting real-time and historical view for the network parameters. Besides, it is used to assemble and store the configured sources of network data and able to present instant comparisons of the current network without any loss of network data [3].

The disadvantage of Watch Point is the visualization will only be understood by network experts.

2) ntop: ntop has been designed for analysing traffic patterns. Some of the system experts have extended ntop by adding embedded NIDS (Network Intrusion Detection System) in order to improve the system. ntop NIDS is very distinctive with its knowledge compare with current NIDS. It is also dynamic and not specified at ntop start-up by means of configuration files [4].

The disadvantages are designed for those network experts and no customization are being allowed in ntop.

3) Nodemap: Nodemap is designed for the purpose to present SNMP queries against network devices as well as to determine the complicated networks link status. The detailed information on network link status will be presented at low-levels visualization together with higher levels summarizations. This is to ensure network computer user can be easily to determine the current state of a network and gained enough information to analyse performance complaints without needing to know every single detail about the network. Besides, Nodemap is also useful for tracking DoS packets flow in complex networks [5].

The disadvantages of this tool are only targeted to network computer user with higher network data knowledge and not permitted for customization from the computer users.

B. Network Security Visualization Tools

1) VISUAL: Visual Information Security Utility for Administration Live (VISUAL) is a network security visualization tool that allows network administrators to examine the communication networks between internal and external hosts, in order to rapidly aware the security conditions of their network [6]. VISUAL applied the concept of dividing network space into a local network address space and a remote network address space (rest of the internet). In order to produce its data visualizations, data will be taken from the log files of Tcpdump or Wireshark. Previously, it was known as Ethereal [7][8] until Summer 2006 due to trademark disagreement. It is an open source tool which contributed to Unix and Windows, especially for network protocol analyser purpose.

The advantage of VISUAL is to provide a quick overview of the current and recent communication patterns among the monitored network. Administrators can specify their network and remote IP by using home and remote IP filter as shown in Figure 2 in [6]. Based on the information provided by IP filter, administrators can identify any single external hosts that are connected with the number of internal hosts from a grid, which may be relevant to be used in their network. The grid represents home hosts; based on connection lines it allows the network administrator to check the total traffic that exchanged between home host and external host [6].

The disadvantages of VISUAL are useful for only small networks such as home network and meaningful for network experts.

2) SCPD: Another network security visualizations tool such as Spinning Cube of Potential Doom (SCPD) is designed for network professional and also presented simple information on the network security frequency and threats extent to beginner [9]. An example of SCPD has been shown in [9].

The advantage of SCPD is that it provided a complete map of internet address space indicating the frequency and origin of scanning activity will be provided by SCPD. User would be able to visualize easily about the sensor data from a large network. Rainbow color map has been used for the cube colors dots of incomplete connections [9]. Port scans on a single host represented by vertical lines and others scan across hosts will be represented by horizontal lines.

The disadvantages of SCPD are simple information is being presented to lower expertise and customization is not provided in this system.

3) PortVis: Another network security visualization tool is PortVis as shown in Figure 1 in [10]. It was focusing on a single host at a time and doing the analysing on it. It designed for outside security specialists.

The main advantage of this tool is to present outside data entities to outside security specialists. Information such as each TCP port during a period of one hour is being visualized and large scale of security occurrence will be detected by PortVis. PortVis also allow for small scale security occurrence detection, which allowed for further investigation.

The drawbacks of PortVis are focusing on a single host at a time and only security specialists will comprehend on the shown information from PortVis.

4) NVIsionIP: Besides that, Figure 1 shown the NVisionIP in [11] is also a visualization tool that targeted to provide and improve the overall situational awareness of the network among network security administrators. A graphical representation of a class-B network and numbers of different views of the data will be presented to network security administrators. There are three main visualization views in a single application of NVisionIP, namely Galaxy, Small Multiple and Machine visualization views. NVisionIP targeted to improve the interactivity among this visualization views by allowing them to transferring data from one visualization views to other visualization views.

The shortcoming of NVisionIP is the information and visualization views only meaningful to security administrators.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Others computer users with lower knowledge will find this view meaningless for them

5) NIVA: Network Intrusion Visualization Application (NIVA) is another network security awareness tool [12]. It is an intrusion detection data visualizer which integrated with haptic features. The novel haptic feature allows users to sense and interactively analyse intrusion detection data over time and also using three-dimensional space.

The advantage of NIVA is it provides visual and other approach for the visual purposes. Users can fully sense the network intrusion by using haptic features.

The disadvantages of NIVA are the approach is working well in individual network instead of huge network and not applicable to beginner or lower network awareness experts.

III. Architecture Of Expert-Aware Approach

A. Two-Dimensional Architecture Development

We proposed an expert-aware approach to designing a system which formulated with a large amount of high-dimensional network data and adaptive for different types of users. Our proposed architecture not only focuses on a small network but also on a complicated network data. In the preliminary phase, we were conducting a knowledge survey among different types of computer users and collecting data from them. This survey is important in order to collect the network knowledge level and requirements on the network from different types of computer users. Diverse computer users provided us with their requirement of network data details. We construct an intelligent expertise classification algorithm which provides a default setting for the expert-aware network data visualization tool based on the knowledge survey results. The system will learn from continual user feedbacks in order to statistically satisfy the needs of majority tool users. Our focus in on network security data and the expert-aware approach looks at the users' expertise level in network security and adapts the most comprehensive visualization screens that are best for the user understanding.

In our initial architecture design, expert levels will be the most crucial and particular component. We will examine the level of computer users. From the experts' examination, we concluded them into initial three different default levels, which are the expert level-one also known as beginner, level-two or intermediate and level-three or advanced. The details of those different levels will be discussed in the following subsections. This subsection will discuss more about the development of two-dimensional screens for expert level-one and level-two whereas the next subsection will discuss more about the development in three-dimensional which targeted expert level-three. The architecture is mostly based on the node concept. A node is an entity (class, in our case with object-oriented programming) containing several elements such as, an icon (type depends on the programming language used), a x coordinates and a y coordinates as an Integer type (to localize the icon in the scene), some Strings containing the different IP addresses, a date type and also a list of nodes.

1) Details of Expert Level-One (Beginner): The expert level-one screen as shown in Figure 1 considers the user as a beginner in computer sciences, or at least someone who has very basic and common computer awareness. Based on the user requirements, system generates the initial screen for computer users, which are Figure 1 and 2. There are three types of data that will be shown on the expert level-one default screen:

a) Node: Composing the network represented by a machine icon, including IP addresses such as IP source, IP destination and date of the analysis, displayed when mouse moving above the concerned node.

b) Address book: Containing every computer shown on the screen, allowing the user to have an overall view of who is connected on the network.

c) Worm detection: The system detects any kind of worms that present in the network and it will immediately launch a pop-up window informing where the infection comes from. An icon will appear on the involved node to show that to the user in a more visual way.

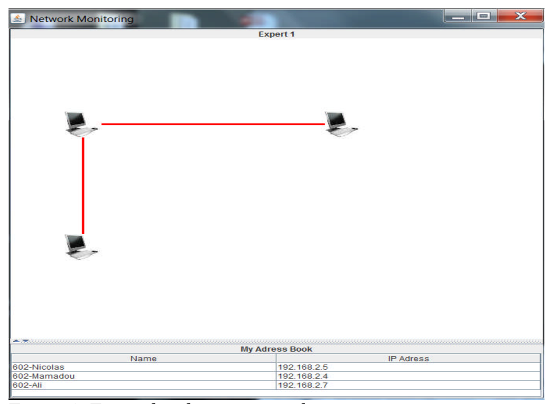


Figure 1. Expert level-one screen shot.

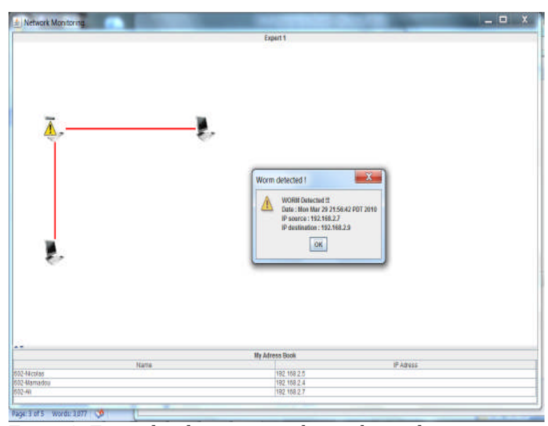


Figure 2. Expert level-one screen shot with simple worm detection alert.

2) Details of Expert Level-Two (Intermediate): Figure 3 showed the screen shot of expert level-two. In this expert level, users consider as someone who has a little knowledge in computer network. Three new types of data have been added to the screen and some interactivity elements have been

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

provided into this expert level. Animation features have been included in the development phase for expert level-two. The links between computers have been replaced by more complex entities exchanges.

a) Packets per sec: This information is represented by the speed of the packets coming from a computer to another. It showed that the packet between the two nodes become faster and the packet per second value of the network become higher.

b) Network utilization: This data is shown using the color of the packets by following this criteria; if it turns out that the network is subject to a high utilization, the color of the packets will be dark. And if the network is very less in used, the color of packets will be slightly lighter.

c) Packets size ratio: It is represented on the screen by the size of the packets that are exchanged between two machines.

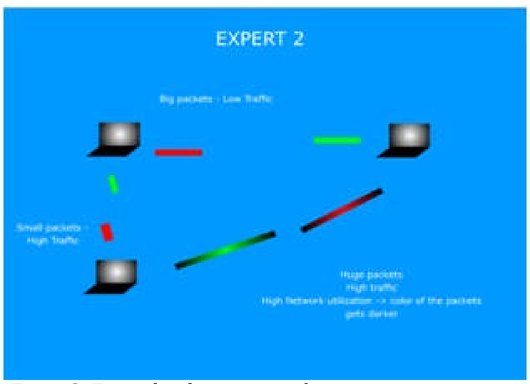


Figure 3. Expert level-two screen shot.

B. Three-Dimensional Architecture Development

Basically the three-dimensional (3D) architecture development is targeted to expert level-three. The computer users with high network knowledge will easily comprehend with the 3D appearances.

There is an EntityNode class to represent a machine (blue sphere) and its IP address. The constructor of this class takes three parameters: the radius of the sphere, the vector locating the sphere and the String which will be display above the machine. Part of the programming has been shown in Figure 4.

The size of the text is then reduced because of the huge default size that Java3D provides to its Text3D instances.

The Request class which make a 3D text going from an EntityNode to another one. The constructor of this class takes three parameters: a first EntityNode, from where the text will come, a second EntityNode which will be the destination of the text. The last parameter is the speed that the request will have to go from the start point to the destination point. Figure 5 has shown the programming to create the text.

Once, we have created the text, we need to use several Java 3D classes to make it move. The most important one is the PositionPathInterpolator object as shown in Figure 6.

1) Details of Expert Level-Three (Advanced): Computer user in expert level-three is expecting to have high awareness on network security data and network data. Figure 7 and 8 shown the screens shot of our initial development which is still ongoing process and will be improved from time to time.

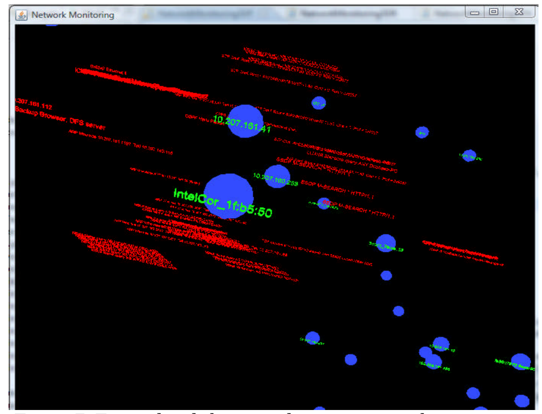


Figure 7. Expert level-three single view screen shot.

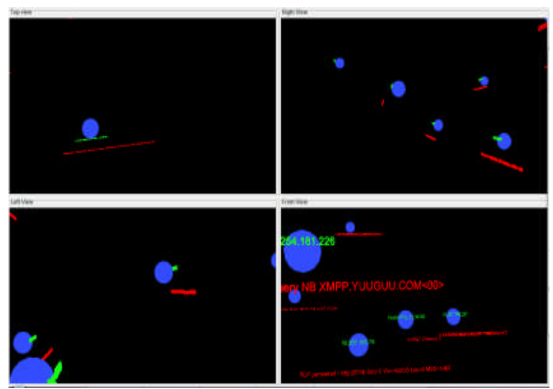


Figure 8. Expert level-three multiple view screen shot.

IV. DISCUSSIONS AND CONTRIBUTIONS

In this section, we will briefly summarize and compare our proposed expert-aware approach with the existing network security visualization tools. A brief comparison summary among existing network data visualization tools according to their advantages and disadvantages is shown in Table 1.

Table 1. Comparison summary between expert-aware approach and existing works.

No.	Tools	Advantages	Disadvantages
1.	Watch Point	1.providing both a real-time and historical view	1. meaningless to beginner user
2.	ntop	1.classifying traffic hence recognizing specific attacks	1. meaningless to beginner user
3.	Nodemap	1.produces visualizations to convey the "holistic" state of the network.	1. meaningless to beginner user

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

4.	VISUAL	1.present a quick overview of the current and recent communication patterns	1. only focus on small network 2. meaningless to beginner user
5.	SCPD	1. present a complete map of internet address space	1. simple information is presented 2. customization not provided
6.	PortVis	1. present outside data entities	1. only focus on single host 2. meaningless to beginner user
7.	NVisionIP	1. present class-B network and numbers of different views of the data	1. meaningless to beginner user
8.	NIVA	1. as an intrusion detection data visualizer which integrated with haptic features	1. working well in individual network 2. meaningless to beginner user
9.	Expert-Aware Approach	1. targeted on different types of computer users 2. focus on small and huge network	1. required input from computer users

Initial results of the implementation of the expert-aware approach for the network data visualization tool show that it is capable of representing several of network data not only on two-dimensional space in a computer but also three-dimensional space. The tool able to represent different level of network data details to different levels of users. Our proposed approach is tested with dataset that has been captured by using network monitoring system and system acceptance surveys have been conducted among diverse computer users (beginner, intermediate and advanced) to get the feedback from them in order to improve the algorithm approach. System features such as effectiveness and efficiency have been improved based on the evaluation analysis result. The visualization effectiveness has been enhanced by presenting sufficient network data to relevant computer user as well as the visualization efficiency has been improved by maximizing network data understanding among computer users.

The results from the evaluation also showed that the expert-aware approach that applied in network data visualization is similar to some other existing network data visualization tools, it lays out complicated network data on comprehensive representation, and added further advantage by making it possible to display very large volume of network data by allowing the different level of computer users to view the different level of network data details. It is able to show not only the small portion of network security data but all relevant data to different types of user.

The main contribution of our approach is targeted to fulfill diverse computer users' requirement on the different levels of network data details. Our approach has also been tested among the researchers and non-researchers from National Advanced IPv6 Centre, Universiti Sains Malaysia.

Besides, small network and complicated network will put in concern in this approach development.

V. Conclusion

In this research, we proposed and implemented an innovative and intuitive expert-aware approach for the network data visualization tools, which improved the existing network data visualization tools. Our experiments in a network lab suggest that the tool can be potential be further improved as the tool has a high potential to a wide range of computer users in the visualization area. The initial result showed that the expert-aware approach has the capability for intelligence adjustment change whenever network data are updated. It will also improve on performance, effectiveness, and efficiency of network data visualization. The well-developed network data visualization approach makes it a promising network data visualization tool for the future.

Acknowledgment

Our special thanks to Institute of Postgraduate Studies (IPS), Universiti Sains Malaysia (USM) for their financial support by awarding Doris Hooi-Ten Wong the Fellowship Scheme. We would like to thank to National Advanced IPv6 (NAv6), Universiti Sains Malaysia (USM) colleagues for their willingness to spare and contribute their guidance.

References

[1] S. M. Bruls, K. Huizing, and J. Van Wijk, "Squarified treemaps," In Proceedings of the Joint Eurographics and IEEE TCVG Symposium on Visualization (VisSym), 33–42, 2000.
[2] M. Allen, P. McLachlan, "NAV Network Analysis Visualization," University of British Columbia, [Online, 29 May 2009].
[3] WildPackets. Watch Point. http://www.wildpackets.com/products/monitoring_and_reporting/watchpoint, [Online, 1 January 2010].
[4] Ntop. http://www.ntop.org/documentation.html, [Online, 1 May 2009].
[5] M. Newton, http://nodemap.internode.on.net/, [Online, 29 May 2009].
[6] R. Ball, G. A. Fink, and C. North, "Home-centric visualization of network traffic for security administration," VizSEC/DMSEC '04: Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security, pages 55–64. ACM Press, 2004.
[7] V. Jacobson, C. Leres, and S. McCanne, TCPdump public repository, http://kb.pert.geant.net/PERTKB/TcpDump, cited September, 2009.
[8] G. Combs, Ethereal downloadable at: http://www.ethereal.com/,cited September, 2009.

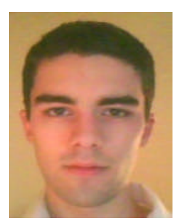
[9] S. Lau, "The Spinning of Potential Doom," Commun. ACM, 47(6):25–26, 2004.
[10] J. McPherson, K. L. Ma, P. Krystosk, Tony Bartoletti, and Marvin Christensen, "Portvis: a tool for port-based detection of security events," In VizSEC/DMSEC '04: Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security, pages 73–81. ACM Press, 2004.
[11] K. Lakkaraju, W. Yurcik, and A. J. Lee. "NVisionIP: Net-flow visualizations of system state for security situational awareness," In VizSEC/DMSEC '04: Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security, pages 65–72. ACM Press, 2004.
[12] K. Nyarko, T. Capers, C. Scott, and K. Ladeji-Osias, "Network intrusion visualization with NIVA, an intrusion detection visual analyzer with haptic integration," in Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2002. HAPTICS 2002 Proceedings, 10th Symposium on, 2002.

AUTHORS PROFILE

Doris Hooi-Ten Wong is a PhD candidate in National Advanced IPv6 Centre (NAv6), Universiti Sains Malaysia (USM). She obtained her B.Sc. (Hons) in Multimedia degree from the Universiti Utara Malaysia in 2008. Her research objectives are to design and develop a new framework, expert-aware approach and intelligence algorithm in network data visualization. She is a member of the Asia-Pacific Advance Network (APAN) as well as the secretariat of APAN Malaysia (APAN-MY).

Kok-Soon Chai (PhD) is a Senior Lecturer of the National Advanced IPv6 Centre (NAv6) at Universiti Sains Malaysia (USM). He was a pioneer and section manager for the embedded software group, Plexus Technology Group in Penang, Malaysia. He led a team of software engineers designing automotive, medical and networking products for US companies. Prior to joining Plexus, he worked at design centers at Agilent and Motorola. He was also involved in research projects sponsored by Airbus UK at the University of Warwick. He is a regular speaker at many conferences. He pioneers the function-class decomposition and UML for embedded software design and presented this approach at the Embedded Systems Conference in Silicon Valley. He obtained a perfect score of 6 out of 6 for the technical content of the presentation averaging from the feedbacks of the attendees. He holds a number of publications in international journal, IEEE conferences, Motorola Software, Systems and Simulation (S3) conference, and a US patent application. He holds a PhD in Engineering from the University of Warwick, UK.

Sureswaran Ramadass (PhD) is a Professor and the Director of the National Advanced IPv6 Centre (NAv6) at Universiti Sains Malaysia (USM). He is also the founder of Mlabs Systems Berhad (MLABS), a public listed company on the MESDAQ. Prof Dr Sureswaran obtained his BsEE/CE (Magna Cum Laude) and Masters in Electrical and Computer Engineering from the University of Miami in 1987 and 1990 respectively. He obtained his doctorate from USM in 2000 while serving as a full time faculty in the School of Computer Sciences. His research areas include the Multimedia Conferencing System, Distributed Systems and Network Entities, Real Time Enterprise Network Monitoring, Real Time Enterprise System Security, Satellite and Wireless Networks, IPv6 Research, Development and Consultancy, and Digital Library Systems.



Nicolas Vavasseur is a Master candidate from Université de Franche Comté. His Master industrial training has been taken in National Advanced IPv6 Centre (NAv6) of year 2010.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

```
//Calculating random coordinates for node 1
float x1 = (float) (Math.random()* 5);
if (k%2!=0) x1 = -x1 ;

float y1 = (float) (Math.random()* 5);
if (k%2==0) y1 = -y1 ;

float z1 = (float) (Math.random()* 5);
if (k%2==0) z1 = -z1 ;

n1 = new NetworkMonitoring3DNodeEntity(0.2f, new Vector3f(x1,y1,z1),element.getIp_source());
node_list.add(n1);
mouseTransform.addChild(n1);
```

Figure 4. Screen shot of programming to create node.

```
// creating the text3D
Text3D textGeom=new Text3D(my_font,new String(request),
        new Point3f((float)(node1.getTranslation_vector().getX()),
                    (float) (node1.getTranslation_vector().getY()),
                    (float)(node1.getTranslation_vector().getZ()))) ;
```

Figure 5. Screen shot of programming to create text.

```
// Creating the animated text
PositionPathInterpolator positionPathInt =
    new PositionPathInterpolator(positionAlpha,
                                 transformGroupTarget,
                                 axe,
                                 noeuds,
                                 positions);
```

Figure 6. Screen shot of programming to create animation.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Load Balancing in Distributed Computer Systems

Ali M. Alakeel
College of Computing and Information Technology
University of Tabuk
Tabuk, Saudi Arabia
Email: alakeel@ut.edu.sa

***Abstract*—Load balancing in distributed computer systems is the process of redistributing the work load among processors in the system to improve system performance. Trying to accomplish this, however, is not an easy task. In recent research and literature, various approaches have been proposed to achieve this goal. Rather than promoting a specific load balancing policy, this paper presents and discusses different orientations toward solving the problem.**

Keywords-distributed systems; load balancing; algorithms; performance evaluation

I. INTRODUCTION

Since the introduction of parallel computers, the main objective has been to allow more than one computer to cooperate in solving the same problem. Obviously, distributing a work load equally between equally capable processors should give the best results. Theoretically speaking, N computers should spend 1/Nth of the time a single computer spends in solving the same problem [1].

Unfortunately, cooperation by itself is not enough and could be potentially disastrous due to such factors as the communication overhead between processors and work imbalance distribution among processors. Having some processors doing more or less work than others will degrade the overall performance of the system and, in the worst case, will cause some the multiprocessors to do nothing at all in reducing the time to solve the problem. Since all processors have to communicate with each other at some point during their computation, a lightly loaded processor will spend most of its time waiting for a subsequent result from an overloaded processor. Consequently, we find that most of the processor's time is spent in waiting rather than doing useful computation as intended.

Load balancing, the process of distributing the work fairly among participating processors, is a sub-problem of a bigger dilemma: distributed scheduling. Distributed scheduling is composed of two parts: local scheduling, which takes care of assigning processing resources to jobs within one node, and global scheduling, which determines which jobs are processed by which processor. Load balancing is a vital ingredient in any acceptable global scheduling policy. The aim of load balancing is to improve system performance by preventing some processors from being overwhelmed with work while others find no work to do.

Achieving the best load balance in a distributed system is not an easy task. A good load balancing policy should consider the following goals: (1) Optimal overall system performance--maximum total processing capacity at acceptable delays; (2) Fairness of service--jobs should be serviced equally regardless of their origin; and (3) Failure tolerance--keeping the system performance at an acceptable level in the presence of partial failures in the system [4].

For the purpose of this presentation we assume the configuration shown in Figure 1. This configuration is a one possible arrangement of a distributed homogenous computer system which has N processors connected through a computer network. In this configuration, processors assumed to be of the same type and have the same power. Tasks (jobs) are individually independent of each other, and can be processed by any processor. Moreover, tasks arrive at each processor P_j has an arrival rate λ_j and a First Come-First-Serve (FCFS) queue discipline is assumed for all queues in the system under consideration.

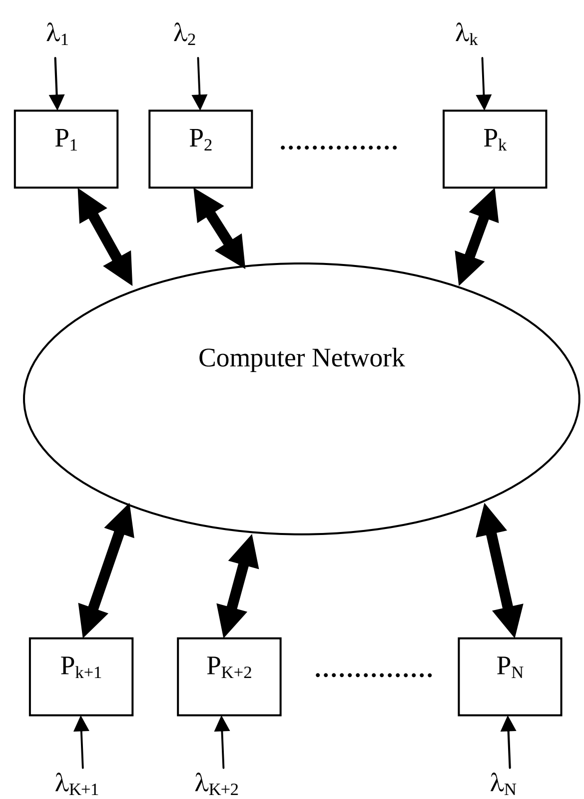


Figure 1. An Example Distributed Computer System Configuration

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Several attempts have been made in the load balancing field, yet with different approaches and orientations. Two different approaches have been taken by researchers in their attempts to achieve load balance in distributed systems: Static and Dynamic.

In the static approach, enough information about the status of all processors in the system is assumed before distributing the work load among them. Once tasks have been assigned to run at specific processors, however, this assignment is final and cannot be changed regardless of any changes occurring later in the system [4], [7], [10], [11], and [18]. Static strategies may be either deterministic or probabilistic. A deterministic strategy assigns tasks to processors based on a fixed criterion, while a probabilistic strategy uses probability values when the assignment is made. For example the load balancing policy that transfers extra tasks from processor A to processor B all the time is deterministic, while the policy that transfer extra tasks from processor A to processor B with probability 0.7 and to processor C with probability 0.3 is a probabilistic one [5]. A static solution's ignorance of system workload fluctuations in decision making is a major disadvantage. On the other hand, static algorithms are easier to work with and analyze [4], [5], and [18]. Several static algorithms have been developed and implemented, some of which can be found in [12]-[14], and [19]-[22].

This paper concentrates almost completely on the dynamic approach, due to its more realistic approach to load balancing. The goal of this paper is not to advance a specific dynamic load balancing policy, but rather to address the problem and present different approaches that have been used to develop a solution for it. Section II presents dynamic load balancing strategies and Section III identifies different methods depicting the assignment of responsibility for conducting load balance. Section IV presents different solutions a load balance strategy could yield, Section V identifies some techniques used to model and analyze a load balance strategy, and Section VI concludes the paper.

II. Dynamic Load Balancing

The dynamic approach looks at the load balancing problem more realistically by assuming little information is available before any assignment is made. It does not presume any knowledge of where a certain task will finally execute or in what environment. Dynamic load balancing algorithms monitor changes on the system work load and redistribute the work load accordingly [4], [7], [10], and [11]. Although the dynamic load balancing approach alleviates the drawbacks of the static approach, it is harder to work with and analyze [4]. Several dynamic algorithms, e.g., [2], [3], [15]-[17], [23]-[26], have been developed and implemented.

Research has shown that the dynamic approach outperforms the static approach and yields better system performance [5]. This section identifies some, not necessarily all or the best, dynamic load balancing strategies that have been reported in literature. It has to be emphasized that the selection of these specific strategies is to serve for presentation purposes only and not to be interpreted as an exhaustive selection or classification. Four strategies will be presented: Bidding, Drafting, Threshold, and Greedy. All presented strategies try to minimize the response time of each process. A comparison of the strategies will be used to demonstrate their relative uses and effectiveness. The bidding strategy is compared with the drafting strategy based on their similarities more than their differences. The way each of them approaches the problem is presented, in addition to identifying the major parameters and properties of each. Communication overhead is used to discriminate between the performances of each strategy.

Similarly, the threshold strategy is compared with the greedy strategy based on their similarities. Since both strategies assume the same amount of information to reach the same objective, their performance will be compared in addition to presenting the features of each strategy.

A. Bidding Strategy

The main concept of this approach is bids. The overloaded processor looking for help in executing some of its tasks requests other processors to submit their bids. Bid information includes the current work load status about each processor. After receiving all bids from participating processors, the original processor selects to whom it will send some of its tasks for execution. A major drawback of this strategy is the possibility that one processor will become overloaded as a result of its winning many bids. To overcome this problem, some variations of this strategy would allow the bidder processor to accept or reject the tasks sent by the original processor. This could be done by allowing it to send a message to the original processor informing it of whether the work has been accepted or rejected. Since a processor's load could change while these messages take place, the final selection might not turn out to be as good as it seems to be at earlier time or vice versa [8]. Different algorithms have been proposed in the literature to determine who gets to initiate the bid, bid information, bid selection, bid participation, and bid evaluation [4], [7]-[11].

The performance of this strategy depends on the amount of information exchanged, the selection of bids, and communication overhead [4]. More information exchange enhances the performance and provides a stronger basis for selection but also requires extra communication overhead [4], [6], and [8]. Examples of this strategy can be found in [27]-[30].

B. Drafting Strategy

The drafting strategy differs from the bidding strategy in the way it allows process migration and in the manner it attempts to achieve load balance. The drafting policy tries to alleviate some of the communication overhead introduced by the bidding strategy. The drafting policy achieves load balance by keeping all processors busy rather than evenly distributing the work load among participating processors (which is one of the objectives of the bidding strategy). In the bidding strategy, to keep all processors evenly loaded, groups of processes will be required to migrate from a heavily loaded processor to a lightly loaded processor. Consequently, it is possible to find some of these processes migrating back as a result of the unpredictable change of the processor's work load. To allow

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

for this problem in the bidding approach, the drafting strategy allows only one process to migrate at a time rather than group migration [8].

The drafting strategy adopts a process migration policy which is based on giving the control to the lightly loaded processors. Lightly loaded processors initiate a process migration instead of having process migration being triggered by a processor being overloaded as in the bidding strategy [8]. In drafting, the number of processes currently at the processor is used for work load evaluation. Each processor maintains its work load and identifies itself as in one of the following states: H-Load, N-Load, or L-load. An H-Load, heavy load, indicates that some of this processor's processes can migrate to other processors. An N-Load, normal load, indicates that there is no intention for process migration. A L-Load, light-load, indicates that this processor is willing to accept some migrant processes. A load table is used at each processor to hold this information about others processors and act as a billboard from which the global information of the system is obtained. When a load change occurs in a processor, it will broadcast its load information to other nodes so that they will update their load tables.

As the processor becomes lightly loaded, i.e. L-Load, it will identify other processors having the status of H-Load from its load table and send them a draft-request message. This message indicates that the drafting processor is willing to accept more work. If by the time it receives this message it is still in H-Load, each remote (drafted) processor will respond by sending a draft-respond message which contains a draft-age information. Otherwise the current load status will be returned to the drafting processor, adopting the concept that a process is allowed to migrate only if it is expecting a better response time and age is associated with each draftable process. Some of the parameters that may be used for age determination are: Process waiting time, process priority, or process arrival [8].

The draft-age is determined by the ages of those processes nominated to be drafted. Various alternatives for draft-age calculations are possible. The selection of the draft age to be the maximum age of all draftable processes, the average age of the draftable processes, or simply the number of draftable processes are some of them [8]. When all draft-response messages are received, the drafting processor calculates draft-standard criteria. Draft-standard criteria are calculated based on the draft-ages received and used to ensure fairness of selection among drafted processes. The choice of draft standard is crucial to the performance of this strategy and is determined at the system design stage. After calculating the draft-standard, a draft-select message is sent to the drafted processor that has the highest draft-age. The drafting processor will send the draft-select message, only if it is still on the L-Load state, otherwise it will not accept any migrating processes.

Research has shown in [8] that the drafting strategy alleviates many drawbacks encountered in the bidding algorithm such as unnecessary communication messages and the possibility of having a bid winner processor being overloaded. Simulation results and detailed comparison results are reported in [8].

C. *Threshold Strategy*

In this type of load balancing algorithms, a threshold value T is used to decide whether a task is executed locally or remotely. The threshold is the processor's queue length. The queue length is the number of processes in service plus the number of processes waiting in the queue. Threshold value assumes static value in an implementation of the algorithm [4] and [5].

In this strategy, a processor will try to execute a newly arriving task locally unless this processor threshold has been reached. In this case, this processor will select another processor randomly and probe it to determine if transferring the task to the probed processor will place it above the threshold or not. If it does not, then the task is transferred to the probed processor which has to execute the task there without any attempts to transfer it to a third one. If it does place it above the threshold, then another processor is selected in the same manner. This operation continues up to a certain limit called the probing limit. After that, if no destination processor is found, the task is executed locally [5]. It should be noted that the threshold strategy requires no exchange of information among processors in order to decide whether to transfer a task or not [5]. This is an advantage because it minimizes communication overhead. Another advantage is that the threshold policy avoids extra transfer of tasks to processors that are above the threshold. It has been shown, in [5], that the threshold algorithm with T=1 yields the best performance for low to moderate system load, and the threshold algorithm with T=2 gives the best performance for high system load.

D. *Greedy Strategy*

In this strategy, the current state of each processor P represented by a function f(n) , where n is the number of tasks currently at the processor. If a task arrives at P and number of tasks n is greater than zero, then this processor looks for a remote processor that has its state less than or equal to f(n). If a remote processor is found with this property, then the task is transferred there. The performance of this strategy depends on the selection of the function f(n). It has been shown in [6] that $f(n) < n$ must holds in order to achieve good performance. Also, n-1, n div 2, n div 3, n div 4, etc. are possible values for f(n). Furthermore, it has been shown in [6] that f(n) = n div 3 yields the best results and that the greedy strategy outperforms the threshold strategy with T=1 in all experiments.

The greedy strategy adopts a cyclic probing mechanism instead of the random selection used in the threshold strategy. In this cyclic probing mechanism, processor i probes processor (i+j) mod N, N representing the number of processors in the system, in the j^{th} probe to locate a suitable destination processor. For example, in a system with 5 processors numbered 0,1,2,3, and 4 respectively, Processor 1 will first probe processor 2. If this attempt is not successful, it will probe 3 and so on. As in the threshold strategy, once a task is transferred to a remote processor it must be executed there [6]. Despite the similarities between the two strategies, it has been demonstrated using simulation results in [6] that the greedy strategy outperforms the threshold strategy. This improvement is attributed to the fact that the greedy strategy attempts to

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

transfer every task that arrives at a busy processor whereas the threshold strategy attempts to transfer only when a task arrives at a processor which has reached the threshold T or higher.

III. RESPONSIBILITY OF LOAD BALANCING

Along with various load balancing strategies which may be applied independently or tailored to enhance the performance of an algorithm for solving a certain problem, different policies of where to put the control of the load balancing algorithm have been proposed in the literature: centralized, distributed, or semi-distributed.

A centralized load balancing strategy assigns a single processor the responsibility of initiating and monitoring the load balance operation. In this strategy, a dedicated processor gathers the global information about the state of the system and assigns tasks to individual processors. Despite its high potential of achieving optimal performance, centralized strategies have some disadvantages: high vulnerability to failures, storage requirements for maintaining the state information - especially for large systems, and the dependability of the performance of the system on the central processor which could result in a bottleneck [9].

In a distributed load balancing strategy, each processor executes the same algorithm and exchanges information with other processors about the state of the system. Each processor may send or receive work on the basis of a sender-initiated or a receiver-initiated policy. In a sender-initiated policy, the sender decides which job gets sent to which receiver. In a receiver-initiated policy, the receiver searches for more work to do. Intuitively, queues are formed at senders if a receiver-initiative policy is used, while they are formed at receivers if a sender-initiative policy is used. Additionally, scheduling decisions are made when a new job arrives at the sender in a sender-initiative, while they are made at the departure of a job in a receiver-initiative policy. The determination of which policy is adopted depends upon the load transfer request which can be initiated by an over-loaded or under-loaded processor. Many distributed strategies belong to either of the two policies. For instance, of the strategies discussed earlier in Sec II, the bidding strategy belongs to the sender-initiated policy, whereas the drafting strategy belongs to the receiver-initiated policy [4], [5], [8], and [9].

It has been demonstrated in [4], [5], [18], using analytical models and simulations, that sender-initiated strategies generally perform better at lower system loads while receiver-initiated strategies perform better at higher system loads, assuming that process migration cost under the two strategies is comparable. Some of the advantages offered by the distributed policy are: Fault tolerance, minimum storage requirements to keep status information, and the availability of system state information at all nodes. The distributed policy still has some disadvantage, one of which is that optimal scheduling decisions are difficult to make because of the rapidly changing environment introduced by the arrivals and departures from individual processors. Another disadvantage is the extra communication overhead is introduced by all processors trying to gather information about each other. To mitigate this overhead, some distributed strategies minimize the amount of information exchanged, which has a negative reflection on the performance of an algorithm.

The semi-distributed policy comes in the middle between centralized and distributed policies. It is introduced to take the best of each and to avoid the major drawbacks of each of the two policies. The semi-distributed strategy is based on the partitioning of the processors into equal sized sets. Each set adopts a centralized policy where a central processor takes charge of load balancing within its set. The sets together adopt a distributed policy where each central processor of each set exchanges information with other central processors of other sets to achieve a global load balance.

It has been shown in [9] that the semi-distributed policy produces a better performance than the centralized and distributed policies. Research demonstrates that each central processor yields optimal load balance locally within its set. Moreover, this policy does not incur high communication overhead while gathering system state information. Although this policy is a mediator between the centralized and the distributed ones, it fits large distributed systems better than small systems.

IV. DIFFERENT OBJECTIVES OF LOAD BALANCING STRATEGIES

Different load balancing strategies have different objectives and yield various solutions. Some solutions are optimal or suboptimal. This section highlights the features of each solution and its relationship with a load balancing policy.

Optimal solutions can be obtained only if complete information regarding the state of the system, as well as the resource needs of a process, is known. An optimal load balance strategy makes optimal assignments based on some criteria function. Examples of optimizing measures are minimizing process completion time, maximizing system throughput [7].

Static load balancing strategies have a higher potential for yielding optimal solutions than dynamic ones. In some situations, however, producing an optimal solution is computationally infeasible. In this case, suboptimal solutions may be targeted. Suboptimal solution could be either approximate or heuristic [7].

An approximate solution uses the same algorithm of producing an optimal solution, but instead of searching the entire solution space, it limits itself to producing a "good" solution in less time rather than a perfect solution. Finding a good approximate solution depends on the availability of a function to evaluate a solution, the time required to evaluate a solution, the ability to judge according to some measurement criteria, the value of an optimal solution, and the availability of mechanism for intelligently reducing the solution space [7].

Heuristic load balancing strategies use static algorithms which make the most realistic assumption regarding the information available about process and system load. Heuristic solutions try to identify parameters that affect system performance indirectly and monitor them. For instance, clustering groups of processes that communicate heavily within

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

one processor would enhance system performance. Although this act affects system performance directly by decreasing the overhead in passing information, it cannot be directly related in a quantitative way to system performance as seen by the user [7]. Four techniques of task allocation algorithms are usually used by static load balancing strategies whether it is trying to produce optimal or approximate solution: solution space enumeration and search, graph theoretic, mathematical programming, or queuing theoretic [7].

V. Load Balancing Algorithm Modeling and Analysis

Before an algorithm is implemented, it is usually analyzed to anticipate the worth of its effectiveness were it to be implemented. Relating to load balancing algorithms, analytical modeling and simulation were the dominating techniques in the literature. They were used extensively to demonstrate and compare various strategies. Queuing theory in particular was used in both analytical and simulation modeling. Analytical modeling could be targeted to analyze the system performance in steady-state or in non-steady-state. Steady-state analysis is based on birth and death Markovian processes while non-steady-state would be based on how the system would perform in the presence of partial failure, i.e., system fault-tolerance is analyzed. Simulation modeling could be either discrete-event or continuous event. In case of discrete-event simulation, simulation languages such as SLAM, SIMAN, GPSS, SIMSCRIPT, etc., or any of high level programming languages such as C or MATLAB, are used to model the system. In the case of continuous modeling, differential and integral equations techniques are used. Steady-state based analysis and discrete-event simulation was heavily used to analyze and model load balance algorithms in the literature.

VI. Conclusion

This paper has attempted to present the most recent ideas and achievements realized in load balancing in distributed systems. The intention has been to provide a suitable understanding of the problem and different approaches that researchers have employed to solve it. Specific load balancing strategies were presented to give an idea of where the research is headed in this field, rather than to elect them over others.

References

[1] H. S. Stone, High-Performance Computer Architecture, 2nd ed. Addison Wesley, Reading, MA, 1990.

[2] S. Dhakal, M. M. Hayat, J.E.Pezoa, C. Yang, and D. Bader, "Dyanmic Load Balancing in Distributed System in the Presence of Delays: A Regeneration-Therory Approach,", IEEE Transactions on Parallel and Distributed Systems, vol. 18, no. 4, April 2007.

[3] L. M. Campos and I. Scherson, "Rate of Change Load Balancing in Distirbuted and Parallel Systesms, " Parallel Computing, vol. 26 no. 9, pp. 1213-1230, July 2000.

[4] Y. Wang and R. Morris, "Load Sharing in Distributed Systems," IEEE Trans. Comput., vol. C-34, no. 3, pp. 204-217, Mar. 1985.

[5] D.L. Eager, E.D. Lazowski, and J. Zahorjan, "Adaptive Load Sharing in Homogeneous Distributed Systems," IEEE Trans. Software Eng., vol. SE-12, no. 5, pp. 662-675, May 1986.

[6] S. Chowdhury, "The Greedy Load Sharing Algorithms," J. Parallel and Distributed Comput., vol. 9, pp. 93-99, May 1990.

[7] T. L. Casavant, "A Taxonomy of Scheduling in General-Purpose Distributed Computing Systems," IEEE Trans. Software Eng., vol 14, no. 2, pp 141-154, February 1988.

[8] L. M. Ni, C. Xu, and T. B. Gendreau, "A Distributed Drafting Algorithm for Load Balancing," IEEE Trans. Software Eng., vol.SE-11, no. 10, pp. 1153-1161, October 1985.

[9] I. Ahmed and A. Ghafoor, "Semi-Distributed Load Balancing for Massively Parallel Multicomputers," IEEE Trans. Software Eng., vol. 17, no. 10, pp 987-1004, October 1991.

[10] K. Ramamritham, J. A. Stankovic, and W. Zhao, "Distributed Scheduling of Tasks with Deadlines and Resource Requirements," IEEE Trans. Comput., vol. 38, no. 8, pp 1110-1123, August 1989.

[11] J. A. Stankovic, K. Ramamritham, and S. Cheng, "Evaluation of a Flexible Task Scheduling Algorithm for Distributed Hard Real-Time Systems," IEEE Trans. Comput., vol. C-34, no. 12, pp. 1130-1143, December 1985.

[12] A. N. Tantawi and D. Tawsley, "Optimal Static Load Balancing in Distributed Computer Systems," J. of Assoc. Comput., vol. 32, no. 2, pp. 445-465, April 1985.

[13] S. H. Bokhari, "Dual Processor Scheduling with Dynamic Reassignment," IEEE Trans. Software Eng., vol. SE-5, no. 4, pp. 341-439, July 1979.

[14] C. Kim and H. Kameda, "An Algorithm for Optimal Static Load Balancing in Distributed Computer Systems," IEEE Trans. Comput., vol. 41, no. 3, pp. 381-384, March 1992.

[15] S. Penmasta and A. T. Chronopoulos, "Dynamic Multi-User Load Balancing in Distributed Systems", 2007 IEEE International Parallel and Distributed Processing Symposium, pp. 1-10, Long Beach, CA, USA, March 2007.

[16] A. Karimi, F. Zarafshan, A. b. Jantan, A. R. Ramli and M. I. Saripan, "A New Fuzzy Approach for Dynamic Load Balancing Algorithm," International Journal of Computer Science and Informaiont Security," vol. 6 no. 1, pp. 001-005 , October 2009.

[17] C.C. Hui and S. T. Chanson, "Improved Strategies for Dynamic Load Balancing," IEEE Concurrency, vol. 7, no. 3, pp. 58-67, July-Sept., 1999.

[18] D. L. Eager and E. D. Lazowski, and J. Zahorjan, "A Comparision of Receiver-Inititated and Sender Initiated Adaptive Load Sharing," Performance Evaluation, 6, pp. 53-68, March, 1986.

[19] J. A. Bannister and K. S. Trivedi, "Task Allocation in Fault-Tolerant Distributed Systems," Acta Inform., vol. 20, pp. 261-281, 1983.

[20] F. Berman and L.Syder, "On Mapping Parallel Algorithms into Parallel Architectures," in 1984 Int. Conf. Parallel Proc., pp. 307-309, August 1984.

[21] X. Tang and S.T. Chanson, "Optimizing Static Job Scheduling in a Network of Hetrogenous Computers," Proc. of the Intl. Conf. on Parallel Processing, pp. 373-382, August 2000.

[22] K. Efe, "Heuristic Models of Task Assignment Scheduling in Distributed Systems," Computer, vol. 15, no. 6, pp. 50-56, June 1982.

[23] G. R. Andrews, D. P. Dobkin, and P.J. Doweny, "Distributed Allocation with Pools of Servers," in ACM SIGACT-SIOPS Symp. Principles of Distributed Computing, pp. 73-83, August 1982.

[24] R. M. Bryant and R. A. Finkel, "A Stable Distributed Scheduling Algorithm," in Proc. 2nd Int. Conf. Dist. Comp., pp. 341-323, April 1981.

[25] T. L. Casavant and J. G. Kuhl, "Design of a Loosely-Coupled Distributed Mutiprocessing Network," in 1984 Int. Conf. Parallel Proc., pp. 42-45, August 1984.

[26] L.M. Ni and K Abani, "Nonpreemptive Load Balancing in A Class of Local Area Networks,", in Proc. Comp. Networking Symp., pp. 113-118, December 1981.

[27] J. A. Stankovic and I. S. Sidhu, "An Adaptive Bidding Algorithm for Processes, Cluster and Distributed Groups," in Proc. 4th Int. Conf. Distributed Compu. Sys., pp. 49-59, 1984.

[28] D. Grosu and A. T. Chronopoulos," Noncooperative Load Balancing in Distributed Systems," Journal of Parallel and Distributed Computing, vol. 65, no. 9, pp. 1022-1034, Sept. 2005.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[29] Z. Zeng and B. Veeravalli, "Rate-based and Queue-based Dynamic Load Balancing Algorithms in Distributed Systems," Proc. of 10th Int. Conf on Prallel and Distributed Systems, pp. 349-356, July 2004.

[30] Z. Khan, R. Singh, J. Alam, and R. Kumar, "Perforamnce Analysis of Dynamic Load Balancing Techniques for Parallel and Distributed Systems," International Journal of Computer and Netwrok Security, vol. 2, no. 2, February 2010.

Ali M. Alakeel (also known as Ali M. Al-Yami) obtained his PhD degree in computer science from Illinois Institute of Technology, Chicago, USA in Dec. 1996, his M.S. degree in computer science from University of Western Michigan, Kalamazoo, USA in Dec. 1992 and his B.Sc. degree in computer science from King Saud University, Riyadh, Saudi Arabia in Dec. 1987. He is currently working as an Assistant Professor at the department of Information Technology, College of Computing and Information Technology, University of Tabuk, Saudi Arabia. His current research interests include automated software testing, distributed computing, cellular networks, and fuzzy logic.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Intrusion Detection using Multi-Stage Neural Network

Sahar Selim, Mohamed Hashem and Taymoor M. Nazmy
Faculty of Computer and Information Science
Ain Shams University
Cairo, Egypt
Sahar.Soussa@gmail.com

Abstract— **Security has become a crucial issue for computer systems. New security failures are discovered everyday and there are a growing number of bad-intentioned people trying to take advantage of such failures. Intrusion detection is a critical process in network security. Intrusion Detection Systems (IDS) aim at protecting networks and computers from malicious network-based or host-based attacks. This paper presents a neural network approach to intrusion detection. We compare the use of our proposed multi-stage to single-stage neural network for intrusion detection using single layer perceptron. The advantage of the proposed mutli-stage system is not only accuracy but also the parallelism as every network can be trained on separate computer which provides less training time. Also the multi-stage powers the system with scalability because if new attacks of specific class are added we don't have to train all the networks but only the branch (the neural networks) affected by the new attack. The results showed that the designed multi-stage network is capable of classifying records with 99.71% accuracy and 98.67% accuracy for single stage network.**

Keywords-component; network intrusion detection; neural network; NSL-KDD dataset

I. INTRODUCTION

The rapid development and expansion of World Wide Web and local network systems have changed the computing world in the last decade. The costs of temporary or permanent damages caused by unauthorized access of the intruders to networks and computer systems have urged different organizations to, increasingly; implement various systems to monitor data flow in their networks. These systems are generally referred to as Intrusion Detection Systems (IDSs) [1].

There exist two main types of network intrusion detection methods: anomaly-based and misuse-based. Misuse detection methods, uses well-defined patterns of the attack that exploit weaknesses in the system and application software to identify the intrusions. A characteristic trait of the intrusion is developed offline, and then loaded in the intrusion database before the system can begin to detect this particular intrusion. It has drawbacks: firstly in most systems, all new attacks will go unnoticed until the system is updated (i.e. they cannot detect new attacks that have never occurred in the training data), creating a window of opportunity for attackers to gain control of the system under attack. Secondly, only known attacks can be detected [2].

Anomaly-based systems (ABS), on the other hand, build statistical models that describe the normal behavior of the network, and flags any behavior that significantly deviates from the norm as an attack. This has the advantage that new attacks will be detected as soon as they take place [3].

II. PREVIOUS WORK

An increasing amount of research has been conducted on the application of neural networks for detecting network intrusions. The idea behind the application of soft computing techniques and particularly ANNs in implementing IDSs is to include an intelligent agent in the system that is capable of disclosing the latent patterns in abnormal and normal connection audit records, and to generalize the patterns to new (and slightly different) connection records of the same class [4].

There are researches implement an IDS using MLP which have the capability of detecting normal and attacks connection as in [5], [6], [7]. They are implemented using MLP of three and four layer neural network. References [8], [4] used three layers MLP (two hidden layers) not only for detecting normal and attacks connection but also identify attack type.

Neural Network was also used for dimension reduction of features as in [9]. The SOM was also applied to perform the clustering of network traffic and to detect attacks in [10], [11], [12] and [13]. In [14], self-organizing maps was used for data clustering and MLP neural networks for detection.

Most of the previous studies that used MLP were implemented with at least three layers. Our study use MLP with no hidden layer to perform less complicated network structure and decrease the computation time. The idea of this study is based on the combination of both ideas which are to be able to identify normal and attack records without exhausting the network of identifying attack type to get higher accuracy and also being able to detect attack type by the next levels. This approach has the advantage to flag for suspicious record even if attack type of this record wasn't identified correctly.

III. DATASET DESCRIPTION

KDDCUP'99 is the mostly widely used data set for the evaluation of these systems. The KDD Cup 1999 uses a version of the data on which the 1998 DARPA Intrusion Detection

Evaluation Program was performed. They set up an environment to acquire raw TCP/IP dump data for a local-area network (LAN) simulating a typical U.S.Air Force LAN.

A. *Types of Networking Attacks*

There are four major categories of networking attacks. Every attack on a network can be placed into one of these groupings [15].

1) Denial of Service Attack (DoS): is an attack in which the attacker makes some computing or memory resource too busy or too full to handle legitimate requests, or denies legitimate users access to a machine. e.g. apache, smurf, Neptune, ping of death, back, mail bomb, UDP storm,etc.

2) User to Root Attack (U2R): is a class of exploit in which the attacker starts out with access to a normal user account on the system (perhaps gained by sniffing passwords, a dictionary attack, or social engineering) and is able to exploit some vulnerability to gain root access to the system. e.g. xlock, guest, xnsnoop, phf, sendmail dictionary etc.

3) Remote to Local Attack (R2L): occurs when an attacker who has the ability to send packets to a machine over a network but who does not have an account on that machine exploits some vulnerability to gain local access as a user of that machine. e.g. perl, xterm.

4) Probing Attack: is an attempt to gather information about a network of computers for the apparent purpose of circumventing its security controls. e.g. satan, saint, portsweep, mscan, nmap etc.

There are some inherent problems in the KDDCUP'99 data set [16], which is widely used as one of the few publicly available data sets for network-based anomaly detection systems. The first important deficiency in the KDD data set is the huge number of redundant records. Analyzing KDD train and test sets, it was found that about 78% and 75% of the records are duplicated in the train and test set, respectively. This large amount of redundant records in the train set will cause learning algorithms to be biased towards the more frequent records, and thus prevent it from learning infrequent records which are usually more harmful to networks such as U2R attacks. The existence of these repeated records in the test set, on the other hand, will cause the evaluation results to be biased by the methods which have better detection rates on the frequent records [15].

The data in the experiment is acquired from the NSL-KDD dataset which consists of selected records of the complete KDD data set and does not suffer from mentioned shortcomings by removing all the repeated records in the entire KDD train and test set, and kept only one copy of each record [15]. Although, the proposed data set still suffers from some of the problems discussed by McHugh [17] and may not be a perfect representative of existing real networks, because of the lack of public data sets for network-based IDSs, but still it can be applied as an effective benchmark data set to help researchers compare different intrusion detection methods. The NSL-KDD dataset is available at [18].

IV. PROPOSED MULTI-STAGE NEURAL NETWORK

A. *Dataset*

In this study we examine using two attacks from each DOS and Probe classes to check the ability of the intrusion detection system to identify attacks from different categories. The sample dataset contains 20000 record for training (10000 normal and 2500 for each attack type) and 1200 for testing (600 normal and 150 for each attack type).

B. *System Architecture*

The proposed system architecture is shown in Fig. 1. The input data are preprocessed. The data must be of uniform representation to be processed by the neural network.

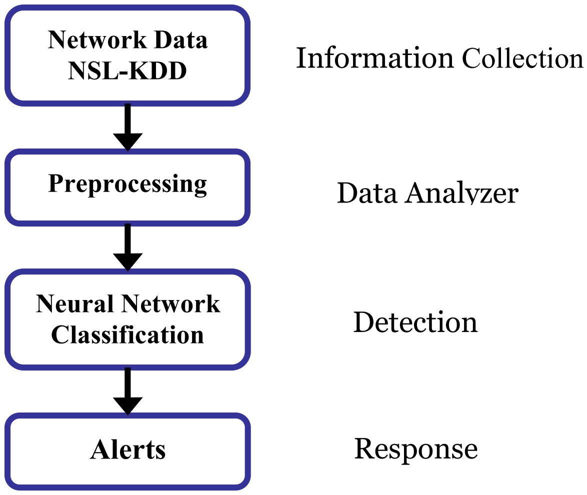


Figure 1. System Architecture.

1) Information Collection: The first module is responsible for data collection. We use the NSL-KDD dataset.

2) Data Analyzer: The second module is for preprocessing.

The preprocessing phase: Features selection, Numerical Representation and Normalization

a) Dimension reduction by excluding the features that are constantly zero over all data records. Hence the data vector is reduced to 30 dimensional vectors.

b) Converts non-numeric features into a standardized numeric representation. This process involved the creation of relational tables for each of the data type and assigning number to each unique type of element. (e.g. protocol_type feature is encoded according to IP protocol field: TCP=0, UDP=1, ICMP=2). This numerical representation was necessary because the feature vector fed to the input of the neural network has to be numerical.

c) It is important to shuffle examples before training so that the network weights are not biased towards a specific attack.

d) The ranges of the features were different and this made them incomparable. Some of the features had binary values where some others had a continuous numerical range

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

(such as duration of connection). As a result, the features were normalized by mapping all the different values for each feature to [0, 1] range.

3) *Detection*

We use neural network for classification. We compare between the proposed multi-stage neural module and single-stage neural network.

a) *Multi-stage Neural Network*

Attacks of the same class have a defined signature which differentiates between attacks of every class/category from others, i.e. DOS attacks have similar characteristics which identifies them from attacks of Probing. That's why there's often misclassification between attacks of the same class. For that reason, we thought of making a multi-stage neural network consisting of three levels as shown in Fig 2:

- **Level 1**: is a Neural Network that identifies attacks from normal
- **Level 2:** is a Neural Network that identifies classes
- **Level 3:** is a neural network that specify attack type

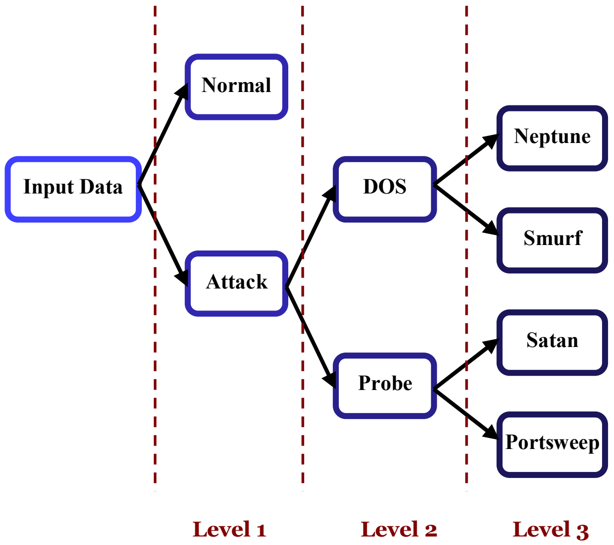


Figure 2. Multi-stage Levels.

The data is input in the first level which identifies if this record is a normal record or attack without exhausting the network to identify the attack name. If the record is identified as an attack then the module would raise a flag to the administrator that the coming record is an attack then the module inputs this record to the second level which identifies the class of the coming attack. If record was classified by network II to be DOS then it would be entered to the DOS network of the third level that identify attacks' type of DOS otherwise it would be introduced to the Probe network. The idea is that if ever the attack name of the third level is misclassified then at least the admin was identified that this record is suspicious after the first level network. Finally the admin would be alerted of the suspected attack type to guide him for the suitable attack response.

1. *Level 1 Architecture:* Neural Network that identifies attacks from normal as shown in Fig. 3.

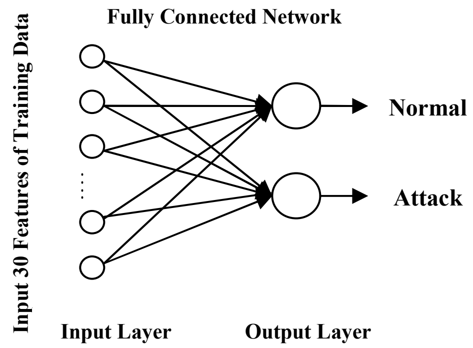


Figure 3. First Level Network which differentiate between Normal and Attack.

2. *Level 2 Architecture:* Neural Network that identifies classes DOS and Probe as shown in Fig. 4.

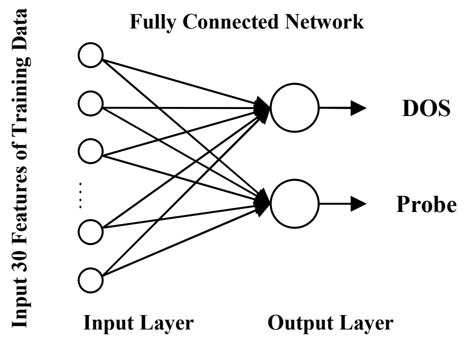


Figure 4. Single Layer Perceptron of Second Level Network which Classify the Attack Class DOS or Probe

3. *Level 3 Architecture:* Neural network that specify attack type

ATTACK TYPE OF DOS CLASS WHETHER NEPTUNE OR SMURF AS SHOWN IN FIG. 5.

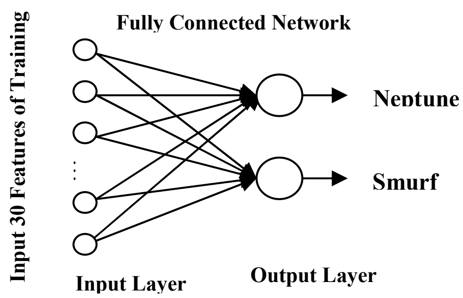


Figure 5. Single Layer Perceptron of third Level Network which Classify Attack type of DOS category.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

ATTACK TYPE OF PROBE CLASS WHETHER SATAN OR PORTSWEEP AS SHOWN IN FIG. 6.

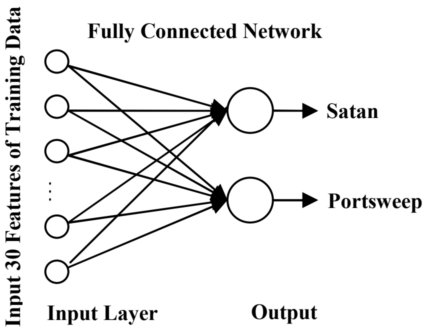


Figure 6. Third Level Network Single Layer Perceptron which Classify Attack type of Probe category.

b) Single Stage Neural Network

In this experiment we examine the use of the neural network for classifying normal and attack type, which means that we input the record and let the MLP identifying the normal and specify the attack name as shown in Fig. 7.

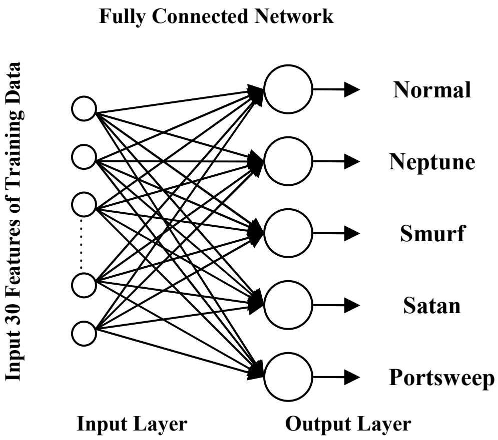


Figure 7. Single-Stage Single Layer Perceptron Network which Classify Normal and Attack type

C. The Over-fitting Problem

One problem that can occur during neural network training is over-fitting. In an over fitted ANN, the error (number of incorrectly classified patterns) on the training set is driven to a very small value, however, when new data is presented, the error is large. In these cases, the ANN has memorized the training examples; however, it has not learnt to generalize the solution to new situations. One possible solution for the over-fitting problem is to find the suitable number of training epochs by trial and error which isn't reasonable for cases that which takes too much time in training. A more reasonable method for improving generalization is called early stopping. In this technique, the available data is divided into three subsets. The first subset is the training set, which is used for training and updating the ANN parameters. The second subset is the validation set. The error on the validation set is monitored during the training process. The validation error will normally decrease during the initial phase of training similar to the training set error. However, when the ANN begins to over-fit the data, the error on the validation set will typically begin to rise. When the validation error increases for a specified number of iterations, the training is stopped, and the weights that produced the minimum error on the validation set are retrieved [19]. In the present study, this training-validation strategy was used in order to maximize the generalization capability of the ANN.

D. Performance Measures

To evaluate our system we used two major indices of performance. We calculate the detection rate and the false alarm rate according to [20] the following assumptions:

- FP: the total number of normal records that are classified as anomalous
- FN: the total number of anomalous records that are classified as normal
- TN: the total number of normal records
- TA: the total number of attack records
- Detection Rate = [(TA-FN) / TA]*100
- False Alarm Rate = [FP/TN]*100

V. EXPERIMENTS AND RESULTS

A. Training of Neural Network

This research aims to examine the difference between a multi-stage MLP and single-stage MLP. Also one of the objectives of the present study is to evaluate the possibility of achieving the same results with this less complicated neural network structure. Using a less complicated neural network is more computationally efficient. Also it would decrease the training time. Therefore we use a single layer perceptron with no hidden layers for all the networks in the two experiments. For each network 20% of the training data were set for cross validation. Early stopping criterion for validation set was applied to stop the training process to prevent over-fitting.

1) Training multi-stage Neural Network

All the 3 levels are a single layer perceptron feed-forward networks (which is the output layer as the input layer contains no processing so it's not considered a layer) with softmax activation function which output results of summation equal to one.

The output layer of first level consists of two neurons one for normal and other for attack. The training process was stopped with mean square error equal 0.0015 at 10000 epochs.

The output layer of second level consists of two neurons one for DOS and other for Probe. The training process was stopped with mean square error equal to 0.000672 at 7914 epochs.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

There are two networks in level three. The first one contains two neurons one for Neptune and the other for smurf. The training process is stopped with mean square error equal to 0.000001 at 1574 epochs.

The second network of level three consists of 2 neurons one for satan and the other for portsweep. The training process was terminated with performance 0.00233 at 5838 epochs.

2) Multi-stage Neural Network Testing Results:

a) Level 1 Testing

The testing phase resulted in success rate 99.83 with error rate 0.167. Table I shows Correct Classification Rate for each of the 2 classes (Attack-Normal) and the total average classification accuracy.

TABLE I. LEVEL 1 CLASSIFICATION RESULTS

Class Name	Training Set	Testing Set
Normal	99.48	99.67
Attack	99.99	100
Average Success Rate	99.74	99.83
Error Rate	0.265	0.167

b) Level 2 Testing

The testing phase resulted in success rate 100. Table II shows the Correct Classification Rate for each of the 2 classes of Level 2 and the total average classification accuracy.

TABLE II. LEVEL 2 CLASSIFICATION RATE

Class Name	Training Set	Testing Set
DOS	99.95	100
Probe	99.77	100
Average Success Rate	99.86	100
Error Rate	0.14	0

c) Level 3 Testing

The testing phase resulted in success rate 99.5 with error rate 0.5. Table III shows the Correct Classification Rate for each of the 4 classes and the total average classification accuracy.

TABLE III. LEVEL 3 CLASSIFICATION RATE

Level 3 Networks	Class Name	Correct Classification	
		Training Set	***Testing Set***
DOS Network	Neptune	100	100
	Smurf	100	100
Probe Network	Satan	100	98.67
	Portsweep	100	99.33
Average Success Rate		100	99.5
Error Rate		0	0.5

3) Training single-stage Neural Network

This network is a single layer feed-forward networks with SoftMax activation. The output layer of this network consists of 5 neurons (normal, Neptune, Smurf, Satan, Portsweep). The training process was terminated with mean square error equal to 0.00034 at 12078 epochs.

4) Single-Stage Neural Network Testing Results

The testing phase resulted in success rate 98.8 with error rate 1.2. Table IV shows the Correct Classification Rate for each of the 5 classes and the total average classification accuracy of the single-stage neural network.

TABLE IV. SINGLE-STAGE CLASSIFICATION RATE

Class Name	Training Set	Testing Set
Normal	99.4	99.33
Neptune	100	99.33
Smurf	99.8	100
Satan	100	100
Portsweep	99.85	94.67
Average Success Rate	99.81	98.67
Error Rate	0.19	1.2

B. Discussion

Building all the networks with a single layer perceptron with no hidden layers gave the advantage of less computation time and less complicated network. The experimental results show that using a multi-stage neural network is more promising than single-stage network as shown in following tables and figures. Table V shows the Correct Classification Rate of testing dataset for each of the 5 classes for both Multi-stage and single-stage.

TABLE V. CLASSIFICATION RATE OF MULTI-STAGE AND SINGLE-STAGE

Class Name	Multi-Stage	Single-Stage
Normal	99.67	99.33
Neptune	100	99.33
Smurf	100	100
Satan	98.67	100
Portsweep	99.33	94.67

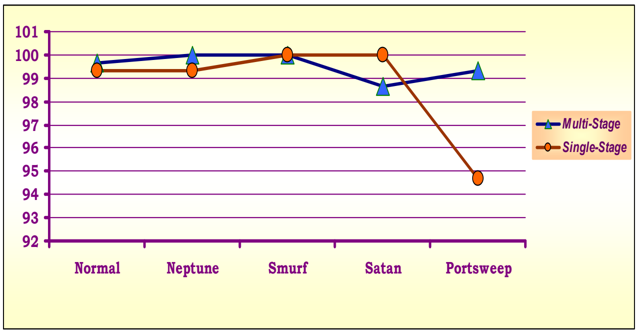


Figure 8. Comparison between Multi-Stage and Single-Stage

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

TABLE VI. FALSE ALARM COMPARISON

Method	Multi-Stage	Single-Stage
FP	2	3
FN	0	7
TN	600	600
TA	600	600
Detection Rate	100	98.83
False Alarm Rate	0.33	0.5

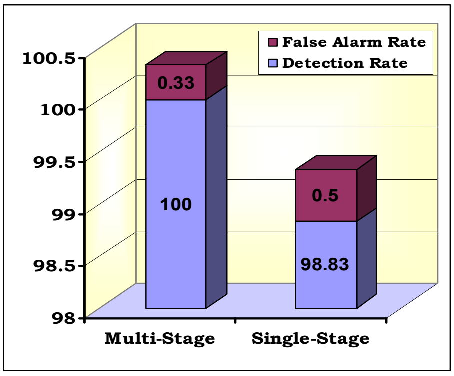


Figure 9. Detection and False Alarm Rate of Multi-Stage and Single-Stage

VI. CONCLUSION AND FUTURE WORK

In this paper we develop a multi-stage neural network and compare its results to results of single-stage neural network. The proposed multi-stage neural network consists of three detection levels. The network data are introduced to the network of the first level which aims to differentiate between normal and attack without exhausting the network in identifying the attack name. If the input record was identified as an attack then the administrator would be alarmed that the coming record is suspicious and then this suspicious record would be introduced to the second level which specifies whether this attack is DOS or probe. The similar characteristics between the attacks of the same class that often results in misclassification between attacks of same class gave the importance of the second level that we have at least identified the class type of the coming attack. The third detection level consists of two networks one to identify attacks of denial of service and the other for probe attacks. Finally the administrator would be alarmed of the expected attack type. The second experiment is for a single stage where the input is classified as one of the 5 classes (normal, Neptune, Smurf, Satan, Portsweep).The results show that the designed multi-stage system has detection rate equal to 100% while the single stage network has detection rate equal to 98.83. The advantage of the proposed mutli-stage system is not only higher accuracy but also the parallelism as every network can be trained on separate computer which provides less training time. Also the multi-stage powers the system with scalability because if new attacks of specific class are added to the dataset we don't have to train all the networks but only the branch (the networks) affected by the new attack.

Future work can include more attack scenarios and use larger dataset. In addition other soft computing techniques will be experimented for classification of U2R and R2L attacks.

REFERENCES

[1] R. A. Kemmerer and G. Vigna, "Intrusion Detection: A Brief Introduction and History," Security & Privacy, IEEE Computer Magazine, pp. 27-30, 2002.

[2] Bolzoni, D., E. Zambon, S. Etalle, and P. Hartel, "POSEIDON: a 2-tier Anomaly-based Network Intrusion Detection System," in Proceedings of the 4th IEEE International Workshop on Information Assurance (IWIA), pp. 144-156, IEEE Computer Society Press, 2006.

[3] D. Bolzoni, S. Etalle, P. Hartel and E. Zambon "POSEIDON: a 2-tier Anomaly-based Network Intrusion Detection System," 2006.

[4] Mohammed Sammany, Marwa Sharawi, Mohammed El-Beltagy, Imane Saroit, "Artificial Neural Networks Architecture For Intrusion Detection Systems and Classification of Attacks," Cairo University, Egypt, 2007.

[5] J.Cannady, "Artificial neural networks for misuse detection," Proceedings of the 1998 National Information Systems Security Conference (NISSC'98), Arlington, VA, pp. 443-456, 1998.

[6] J. Ryan, M. Lin, and R. Miikkulainen, "Intrusion Detection with Neural Networks," AI Approaches to Fraud Detection and Risk Management: Papers from the 1997 AAAI Workshop, Providence, RI, pp. 72-79, 1997.

[7] Srinivas Mukkamala, "Intrusion detection using neural networks and support vector machine," Proceedings of the 2002 IEEE International Honolulu, HI, 2002.

[8] M. Moradi, and M. Zulkernine, "A Neural Network Based System for Intrusion Detection and Classification of Attacks," IEEE International Conference on Advances in Intelligent Systems - Theory and Applications, Luxembourg-Kirchberg, Luxembourg, November 15-18, 2004.

[9] Y. Bouzida, F.e.e. Cuppens, N. Cuppens-Boulahia, S. Gombault, "Efficient intrusion detection using principal component analysis," in: Proceedings of the 3ème Conférence sur la Sécurité et Architectures Réseaux (SAR), Orlando, FL, USA, 2004.

[10] L. Girardin, "An eye on network intruder-administrator shootouts," In Proceedings of the Workshop on Intrusion Detection and Network Monitoring (ID'99), pages 19–28, Berkeley, CA, USA, 1999. USENIX Association.

[11] M. Ramadas, S. Ostermann, and B. Tjaden, "Detecting anomalous network traffic with self-organizing maps," In Recent Advances in Intrusion Detection, 6th International Symposium, RAID 2003, pages 36–54, 2003.

[12] S. Zanero, "Improving Self Organizing Map Performance for Network Intrusion Detection," International Workshop on Clustering High-Dimensional data and its applications, SDM 05 SIAM conference On Data Mining, page. 30-37, 2005.

[13] P. Lichodzijewski, A. N. Zincir-Heywood, M. I. Heywood, "Dynamic intrusion detection using self-organizing maps," Proceedings of the 14th Annual CITASS, Ottawa, Canada, May 2002.

[14] A. Bivens, C. Palagiri, R. Smith, B. Szymanski and M. Emrechts, "Network-Based Intrusion Detection Using Neural Networks," Intelligent Engineering Systems through Artificial Neural Networks, Vol. 12, Proc. ANNIE, 2002.

[15] M. Tavallaee, E. Bagheri, W. Lu, and A. Ghorbani, "A Detailed Analysis of the KDD CUP 99 Data Set," Submitted to Second IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA), 2009.

[16] KDD Cup 1999. Available on: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html, October 2007

[17] J. McHugh, "Testing intrusion detection systems: a critique of the 1998 and 1999 darpa intrusion detection system evaluations as performed by lincoln laboratory," ACM Transactions on Information and System Security, vol. 3, no. 4, pp. 262–294, 2000.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[18] "Nsl-kdd data set for network-based intrusion detection systems." Available on: http://nsl.cs.unb.ca/NSL-KDD/, March 2009

[19] MATLAB online support: www.mathworks.com/access/helpdesk/help/techdoc/matlab.shtml

[20] S.T. Sarasamma, Q.A. Zhu, and J. Huff, "Hierarchal Kohonenen Net for Anomaly Detection in Network Security," IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics, 35(2), 2005, pp. 302-312.

AUTHORS PROFILE

Sahar Selim Fouad Bachelor of Computer Science, Faculty of Computer & Information Science, Ain Shams University. Currently working for master degree. Fields of interest are intrusion detection, computer and networks security.

Mohamed Hashem Abdel-Aziz Professor in computer science, Ain Shams University. Currently head of information systems department, faculty of Computer and Information Science, Ain Shams University. Fields of interest are computer networks, Ad-hoc and wireless networks, Qos Routing of wired and wireless networks, Modeling and simulation of computer networks, VANET and computer and network security.

Taymoor Mohammed Nazmy Professor in computer science, Ain Shams University. He served before in faculties of Sciences, and education as a lecturer for over 12 year. He was the director of the university information network. Currently vice dean of higher studies and researches, faculty of Computer and Information Science, since 2007. Fields of interest are image processing, pattern recognition, artificial neural networks, networks security and speech signal analysis.

Improvement of the Performance of Advanced Local Area Optical Communication Networks by Reduction the Effects of the Propagation Problems

Mahmoud I. Abd-Alla ***Fatma M. Aref M. Houssien**

Electronics & Communication Department, Faculty of Engineering, Zagazig University, EGYPT

Email: mabdalla@gmail.com, *E-mail: fatma_shahin2010@yahoo.com

Abstract

In the present paper, the improvement of transmission distance and bit rates of Advanced Local Area Optical Communication Networks (ALAOCN) are investigated by reducing the effects of propagation problems over wide range of the affecting parameters. Dispersion characteristics in high-speed optical transmission systems are deeply analyzed over a span of optical wavelengths from 1.2 μm up to 1.65 μm. Two different fiber structures for dispersion management are investigated. Two types of fabrication material link of single mode fiber made of Germania doped Silica and plastic fibers are suggested. Two successive segments of single mode fiber made of Germania doped Silica are suggested to be employed periodically in the long-haul systems. The two successive segments are: i) of different chemical structures (x), or ii) of different relative refractive index differences (Δn). As well as the total spectral losses of both fabrication materials and total insertion losses of connectors and splices of these fabrication materials are presented under the thermal effect of to be processed to handle both transmission lengths and bit rates per channel for cables of multi links over wide range of the affecting parameters. Within soliton and maximum time division multiplexing (MTDM) transmission techniques, both the transmission bit rate and capacity-distance product per channels for both of silica doped and plastic fabrication materials are estimated. The bit rates are studied within thermal sensitivity effects, loss and dispersion sensitivity effects of the refractive index of the fabrication core materials are taken into account to present the effects on the performance of optical fiber cable links. Dispersion characteristics and dispersion management are deeply studied where two types of optical fiber cable core materials are used. A chromatic dispersion management technique in optical single mode fiber is introduced which is suitable for (ALAOCN) to facilitate the design of the highest and the best transmission performance of bit rates in optical networks.

Keywords: Propagation problems, Single mode fiber (SMF), Fiber losses, Dispersion types, Dispersion management, Soliton Bit rate thermal sensitivity, optical link design, Thermal effects, Advanced-optical networks.

1. Introduction

Fiber optic transmission and communication are technologies that are constantly growing and becoming more modernized and increasingly being used in the modern day industries [1]. Dispersion occurs when the light traveling down a fiber optic cable "spreads out," becomes longer in wavelength and eventually dissipates. Attenuation, a reduction in the transmitted power, has long been a problem for the fiber optics community. However, researchers have established three main sources of this loss: absorption, scattering, and dispersion [2, 3]. Fiber to the Home (FTTH) technology is one of the main research objectives of the lasts years in optical fiber communication. The increasing development of data communications and the emerging of applications demand a redesign of the access networks in order to accomplish new bandwidth and latency requirements. Wireless communications are a good alternative for quick deployments and low cost implementations but this technology cannot compete against optical communications in terms of available bandwidth, latency and robustness. Since 1980, several techniques have been proposed and applied to reduce such phenomenon which severely reduces the transmitted bit-rate [4, 5]. The rapid increase of transmission capacity need is requiring higher speed optical communication system. However, the upgrade of most installed system at third window to multi-Giga-bit rate is limited by the high linear chromatic dispersion of the optical standard fiber deployed worldwide [6, 7]. To upgrade existing networks based on standard single-mode 1310 nm optimized optical fibers, several all-optical dispersion compensation techniques have been proposed [6].Recent progress in optical fiber amplifier technology makes fiber dispersion the ultimate limiting factor for high-speed long-distance optical fiber transmission. Low-chirp, high-speed optical sources are indispensable for long-haul multi Giga bit-per-second optical communication systems [7].Traffic demand has been increasing steadily in the last few years. In order to support this increasing traffic demand the optical links between the main cities, which are typically terrestrial links with hundreds of kilometers operating at 10 Gbit/sec per channel, have to be upgraded. A solution for the upgrading of these links is to increase the bit rate per channel to 40, 80 or even to 160 Gbit/s. Access optical networks are capable of solving those

requirements for present and future applications. The recent explosive growth of the internet has triggered the introduction of broadband access network based on FTTH. To deal with various demands [6], access and metro networks require scalability in terms of capacity and accommodation and flexibility with regard to physical topology [7]. Therefore, new specific components are required. The advanced high speed technology of core networks is expected to provide cost effective migration of the component solutions towards access applications; however, improvements in terms of the integration of functions and low cost packaging have to be made. Dense wavelength division multiplexing (DWDM) is widely becoming accepted as a technology for meeting growing bandwidth, and WDM systems beginning to be deployed in undersea telecommunications links [8, 9].

As complexity of optical dense wavelength division multiplexing (DWDM) networks increases due to the large number of channels involved, managing the large spectral variations in the dispersion and gain becomes more difficult as the desired spectral bandwidth increases. Dispersion managed soliton, now being developed by a number of different groups, can resolve the technical problems that in the past have prevented the use of the soliton transmission format in optical fiber communication systems [10-12]. Optical solitons are stable nonlinear pulses formed in optical fibers when the nonlinearity induced by the optical intensity is sufficient to balance the dispersion of the fiber. In an ideal lossless fiber, solitons would not distort in either the time or frequency domains, regardless of the distance over which they propagated. A dispersion managed fiber is made by alternating sections of positive and negative dispersion fiber to create a transmission line with high local dispersion and low total dispersion [13].

2. Modeling Basics and Analysis

Special emphasis is given to the propagation problems in silica-doped and plastic fibers as promise links in long and short–distance advanced optical communication networks. Silica-doped and plastic fibers characteristics (spectral loss and chromatic dispersion) are thermal dependent, thus, these two variables must be taken into account when studying the transmission capacity of the fibers. The processed propagation problems in this study will deeply defined, analyzed, investigated parametrically and treated over wide range of the affecting parameters.

2.1. Simplified attenuation model

2.1.1. Silica-doped fibers attenuation model

Based on the models are given in reference [14], the spectral losses of silica-doped fibers are cast as:

$$\alpha = \alpha_I + \alpha_S + \alpha_{UV} + \alpha_{IR} \text{ dB/km} \quad (1)$$

Where $\alpha_I \equiv \text{the intrinsic loss} \cong 0.003$ dB/km, and (2)

$$\alpha_S \equiv \text{Rayleigh scattering} = \left(\frac{0.75 + 66\ \Delta n}{\lambda^4}\right)\left(\frac{T}{T_0}\right) \text{ dB/km} \quad (3)$$

Where we have assumed that the scattering loss is linearity is related to the ambient temperature (T) and (T_0) is a reference temperature (300 °K), (Δn) and (λ) are the relative refractive index difference and optical signal wavelength respectively. The absorption losses (α_{UV}: Ultra-violet losses) and (α_{IR}: Infra-red losses) are given as in reference [14]:

$$\alpha_{UV} = 1.1\times10^{-4}\, \omega_{ge} \% \; e^{4.9\lambda} \text{ dB/km} \quad (4)$$

$$\alpha_{IR} = \left(7\times10^{-5}\; e^{-24/\lambda}\right)^2 \text{ dB/km} \quad (5)$$

Where (ω_{ge} %) is the weight percentage of Germania, GeO_2 added to optical silica fibers to improve its optical transmission characteristics.

2.1.2. Plastic fibers attenuation model

Plastics PMMA Polymethyl Methacrylate, as all any organic materials, absorb light in the ultraviolet spectrum region. The mechanism for the absorption depends on the electronic transitions between energy levels in molecular bonds of the material. Generally the electronic transition absorption peaks appear at wavelengths in the ultraviolet region, and their absorption tails have an influence on the plastic optical fiber (POF) transmission loss [15]. According to urbach's rule, the attenuation coefficient α_e due to electronic transitions in POF is given by the following expression [15]:

$$\alpha_e(PMMA) = 1.10\times10^{-5}\ \exp\left(\frac{8}{\lambda}\right) \text{ dB/km} \quad (6)$$

Where: (λ) is the optical signal wavelength of light in (μm). In addition, there is another type of intrinsic loss, caused by fluctuations in the density, and composition of the material, which is known as Rayleigh scattering. This phenomenon gives the rise to scattering coefficient (α_R) that is inversely proportional to the fourth power of the wavelength, i.e., the shorter is (λ) the higher the losses are. For POF, it is shown that (α_R) is [16]:

$$\alpha_R(PMMA) = 13\left(\frac{0.633}{\lambda}\right)^4 \text{ dB/km} \quad (7)$$

Then the total losses of plastic optical fibers are given:

$$\alpha_{total}(PMMA) = 1.10\times10^{-5}\exp\left(\frac{8}{\lambda}\right) + 13\left(\frac{0.633}{\lambda}\right)^4 \text{ dB/km} \quad (8)$$

2.1.3. Connector and splice attenuation model

There are many types of connectors developed for fiber cable. A connector is used to join a fiber cable to a transmitter or receiver, or is used to join together strands of fiber. A connector for fiber is similar in concept to a traditional electrical connector, but the fiber connector is actually more delicate, as it must precisely align the internal fibers to insure a proper flow of data through the cables. Before connecting one fiber with the other fiber in the fiber optic communication link, one must decide whether the joint should be permanent or demountable.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Based on this, two types of joints are presented. A permanent joint is done by splice and a demountable joint is done by connector. The insertion loss of any connector can be expressed as given by [17]:

$$IL(Insertion\ Loss) = 10\log\left(\frac{P_i}{P_t}\right) \text{ dB/km} \qquad (9)$$

Where (P_i) is the incident power in (mWatt) and (P_t) is the transmitted power in (mwatt). For single mode fibers (SMF), the Fresnel reflection loss caused by the differences between the refractive indices of the silica-doped, ($n=n_1$) and plastic fibers, ($n=n_2$) and the material separation are given as the following [18]:

$$Loss = 20\log\left(\frac{4n_1 n_{clad}}{(n_1+n_{clad})^2}\right) \text{ dB/km} \qquad (10)$$

$$Loss = 20\log\left(\frac{4n_2 n_{clad}}{(n_2+n_{clad})^2}\right) \text{ dB/km} \qquad (11)$$

Where (n_1) is the refractive-index of silica-doped core material, and (n_2) is the refractive-index of plastic core material. The cladding refractive-index can be expressed as a function of both silica-doped and plastic core refractive-indices and relative refractive-index difference as the following:

$$n_{clad} = (1-\Delta n)n_1 \qquad (12)$$

$$n_{clad} = (1-\Delta n)n_2 \qquad (13)$$

Then by substituting with equations (12, 13) into equations (10, 11), we can obtain:

$$Loss\,(silica-doped) = 20\log\left(\frac{4(1-\Delta n)n_1^2}{(2n_1-\Delta n\cdot n_1)^2}\right) \text{ dB/km} \qquad (14)$$

$$Loss\,(plastic) = 20\log\left(\frac{4(1-\Delta n)n_2^2}{(2n_2-\Delta n\cdot n_2)^2}\right) \text{ dB/km} \qquad (15)$$

2.2. Simplified dispersion model analysis

2.2.1. Silica-doped fiber dispersion model

We have employed Germania-doped Silica fiber as a communication channel, where (x) is the mole fraction of Germania added to silica material. The refractive index of silica-doped material may be evaluated following the three terms Sellmeier equation [19]:

$$n_1^2 = 1 + \sum_{i=1}^{3}\frac{A_i\lambda^2}{\lambda^2-\lambda_i^2} \qquad (16)$$

Where (n_1), (λ), (A_i), and (λ_i) are the core refractive index, the oscillator strength and the oscillator wave length respectively. For GeO_2-SiO_2 fibers of x % GeO_2 (mole), the two sets (A_i) and (λ_i) are given [19]:

$$A_1 = (0.6961663 + 0.11070010x)\,f_{T1}, \qquad (17)$$

$$A_2 = (0.4079426 + 0.31021588x)\,f_{T1}, \qquad (18)$$

$$A_3 = (0.8974794 - 0.04331091x)\,f_{T1}, \qquad (19)$$

$$\lambda_1 = (0.0684043 + 0.00568306x)\,f_{T2}, \qquad (20)$$

$$\lambda_2 = (0.1162414 + 0.03772465x)\,f_{T2}, \qquad (21)$$

And $\lambda_3 = (9.896161 + 1.94577x)\,f_{T2}$. (22)

Where $f_{T1}=0.93721+2.0857\text{x}10^{-4}T$, $f_{T2}=T_0/T$, (T_0) is a reference temperature and T is the medium (fiber temperature) [19-21]. Chromatic dispersion (a cause of pulse spreading) arises from the use of sources with a finite spectral spread, since signals impressed on a fiber at different wavelengths will have different group velocities. The transit time $\tau(\lambda)$ of a mode at a wavelength (λ) may be related to that at the mean source wavelength (λ_o) by expanding as a Taylor series about (λ_o) as given in [19-21]:

$$\tau(\lambda) = \tau(\lambda_o) + (\lambda-\lambda_o)\left.\frac{d\tau}{d\lambda}\right|_{\lambda_o} + \frac{1}{2}(\lambda-\lambda_o)^2\left.\frac{d^2\tau}{d\lambda^2}\right|_{\lambda_o}$$

$$+\frac{1}{6}(\lambda-\lambda_o)^3\left.\frac{d^3\tau}{d\lambda^3}\right|_{\lambda_o} \qquad (23)$$

$$\Delta\tau = \Delta\lambda\left.\frac{d\tau}{d\lambda}\right|_{\lambda_o} + \frac{1}{2}(\Delta\lambda)^2\left.\frac{d^2\tau}{d\lambda^2}\right|_{\lambda_o} + \frac{1}{6}(\Delta\lambda)^3\left.\frac{d^3\tau}{d\lambda^3}\right|_{\lambda_o} \qquad (24)$$

Noting that: $\tau = \frac{Lm}{C}$ (25)

Where (L), (C), and (m) are the fiber length, the velocity of light in vacuum, and the group index for the mode respectively, (m) is given by assuming good confinement in [20]: $m = n_1 - \lambda\frac{dn_1}{d\lambda}$ (26)

We can deduce that

$$\tau = \frac{L}{C}\left[n_1 - \lambda\frac{dn_1}{d\lambda}\right] \qquad (27)$$

The use of Eq. (27) into Eq. (24) yields per unit length:

$$\Delta\tau = \frac{\Delta\lambda}{C}.\lambda\left.\frac{d^2n_1}{d\lambda^2}\right|_{\lambda_o} - \frac{(\Delta\lambda)^2}{2C}\left[\lambda\frac{d^3n_1}{d\lambda^3} + \frac{d^2n_1}{d\lambda^2}\right]\Bigg|_{\lambda_o}$$

$$-\frac{(\Delta\lambda)^3}{6C}\left[\lambda\frac{d^4n_1}{d\lambda^4} + 2\frac{d^3n_1}{d\lambda^3}\right]\Bigg|_{\lambda_o} \qquad (28)$$

Where higher-order dispersion modes are considered, following the same spirit of Refs. [20, 21], in separating the various contributions to the total chromatic dispersion in single mode fibers of radius (a), we have:

($\Delta\tau_{ch}/\Delta\lambda.L$) = D_t = total chromatic dispersion coefficient= (M_{md} + M_{wd} + M_{pd}), (29)

Where: M_{md} = material dispersion coefficient

$$= \frac{\lambda}{C}\left.\frac{d^2n_1}{d\lambda^2}\right|_{\lambda_o} + \frac{\Delta\lambda}{2C}\left[\lambda\frac{d^3n_1}{d\lambda^3} + \left.\frac{d^2n_1}{d\lambda^2}\right|_{\lambda_o}\right]$$

$$-\frac{(\Delta\lambda)^2}{6C}\left[\lambda\frac{d^4n_1}{d\lambda^4} + 2\left.\frac{d^3n_1}{d\lambda^3}\right|_{\lambda_o}\right] \qquad (30)$$

M_{wd} = waveguide dispersion coefficient

$$= \frac{n_1}{C}.\frac{\Delta n}{\lambda}\left(\frac{m}{n_1}\right)^2 M(V) \qquad (31)$$

M_{pd} = profile dispersion coefficient

$$= \frac{n_1\Delta n'}{C}\left(\frac{\lambda}{4}\frac{\Delta n'}{\Delta n} - \frac{m}{n_1}\right)M(V) \qquad (32)$$

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

$$\Delta n = \frac{(n_1 - n_{clad})}{n_1} \quad (33)$$

$$\Delta n^{\backslash} = \frac{d\Delta n}{d\lambda} \quad (34)$$

$$V = 2\pi a \sqrt{n_1^2 - n_{clad}^2} / \lambda \text{, and} \quad (35)$$

$$M(V) = 0.08 + 0.549(2.834 - V)^2 \quad (36)$$

Equation (30), in our suggested basic model, accounts for the material dispersion (the first, the second, and the third -order dispersion effects simultaneously). Then the use of Eq. (16) yields:

$$n_1 n_1^{\backslash} = \sum_{i=1}^{3} \frac{-A_i \lambda_i^2 \lambda}{\left(\lambda^2 - \lambda_i^2\right)^2} \quad (37)$$

$$n_1 n_1^{\backslash\backslash} + n_1^{\backslash 2} = \sum_{i=1}^{3} \frac{A_i \lambda_i^2 (3\lambda^3 + \lambda_i^2)}{\left(\lambda^2 - \lambda_i^2\right)^3} \text{, and} \quad (38)$$

$$n_1 n_1^{\backslash\backslash\backslash} + 3n_1^{\backslash} n_1^{\backslash\backslash} = \sum_{i=1}^{3} \frac{-12 A_i \lambda_i^2 \lambda (\lambda^2 + \lambda_i^2)}{\left(\lambda^2 - \lambda_i^2\right)^4} \quad (39)$$

$$n_1 n_1^{\backslash\backslash\backslash\backslash} + 4n_1^{\backslash} n_1^{\backslash\backslash\backslash} + 3n_1^{\backslash\backslash 2} = \sum_{i=1}^{3} \frac{12 A_{iT} \lambda_{iT}^2 (5\lambda^4 + 10\lambda_{iT}^2 \lambda^2 + \lambda_{iT}^2)}{\left(\lambda^2 - \lambda_{iT}^2\right)^5} \quad (40)$$

2.2.2. Plastic fiber dispersion model

The plastic cable core material which the investigation of the spectral variations of the refractive-index (n_2) requires Sellemeier equation given in [22]:

$$n_2^2 = 1 + \sum_{i=1}^{3} \frac{B_i \lambda^2}{\lambda^2 - \lambda_i^2} \quad (41)$$

For the plastic fiber material, the coefficients of the Sellmeier equation and refractive-index variation with ambient temperature are [22]: B_1= 0.4963, λ_1= 0.6965 (T/T_0), B_2= 0.3223, λ_2= 0.718 (T/T_0), B_3= 0.1174, and λ_3= 9.237. Then the first and second differentiation of Eq. (41) with respect to (λ) yields:

$$n_2 n_2^{\backslash} = \sum_{i=1}^{3} \frac{-B_i \lambda_i^2 \lambda}{\left(\lambda^2 - \lambda_i^2\right)^2} \quad (42)$$

$$n_2 n_2^{\backslash\backslash} + n_2^{\backslash 2} = \sum_{i=1}^{3} \frac{B_i \lambda_i^2 (3\lambda^3 + \lambda_i^2)}{\left(\lambda^2 - \lambda_i^2\right)^3} \quad (43)$$

The total chromatic dispersion (D_t) of the output pulse width in single mode fibers of (POF) is given by [22]:

$$D_t = \frac{\Delta \tau}{\Delta \lambda \cdot L} = \left(W_{md} + P\right) \text{ nsec/nm.km} \quad (44)$$

The output pulse width from single mode plastic optical fiber (POF) was taken into account both material and profile dispersions, and thus modal dispersion is equal to zero for single mode fibers [23]:

$$W_{md} = \left(-\frac{\lambda^3}{c} \frac{dn_2}{d\lambda} - \frac{2\lambda}{c} \left(\frac{d^2 n_2}{d\lambda^2} \right) (N_1 \Delta n) \times C_1 \left(\frac{2\alpha}{\alpha + 2} \right) \right)^{1/2} \quad (45)$$

$$P = \left(\left(\frac{N_1 \Delta n}{c\lambda} \right)^2 \left(\frac{\alpha - 2 - \varepsilon}{\alpha + 2} \right)^2 \times \frac{2\alpha}{3\alpha + 2} \right)^{1/2} \quad (46)$$

Where the group index for the mode is given by:

$$N_1 = n_2 - \lambda \frac{dn_2}{d\lambda} \quad (47)$$

Where: ($\Delta\tau$) is the total pulse spreading due to chromatic dispersion in nsec, (W_{md}) is the material dispersion coefficient in nsec/nm.km, (P) is the profile dispersion coefficient in nsec/nm.km, and (Δn) is the relative refractive index difference defined as:

$$\Delta n = \frac{n_2^2 - n_{cladding}^2}{2 n_2^2} \quad (48)$$

Where (n_2) is the cable core refractive index, ($n_{cladding}$) is the core cladding refractive index, and (C_1) is a constant and is given by the following expression:

$$C_1 = \frac{\alpha - 2 - \varepsilon}{\alpha + 2} \quad (49)$$

Where (α) is the index exponent, (ε) is the profile dispersion parameter, and is given by the following:

$$\varepsilon = -\frac{2n_2}{N_1} \frac{\lambda}{\Delta n} \frac{d\Delta n}{d\lambda} \quad (50)$$

3. Transmission Techniques for Reducing Propagation Problems

The need of communication is an all time need of human beings. For communication some channel is needed. Fiber is one channel among many other channels for communication. The dispersion phenomenon is a problem for high bit rate and long haul optical communication systems. An easy solution of this problem is optical solitons pulses that preserve their shape over long distances. Soliton based optical communication systems can be used over distances of several thousands of kilometers with huge information carrying capacity by using optical amplifiers. Soliton communication systems are a leading candidate for long-haul light wave transmission links because they offer the possibility of dynamic balance between group velocity dispersion (GVD) and self-phase modulation (SPM), the two effects that severely limit the performance of non soliton systems. Most system experiments employ the technique of lumped amplification and place fiber amplifiers periodically along the transmission line for compensating the fiber loss. However, lumped amplification introduces large peak-power variations, which limit the amplifier spacing to a fraction of the dispersion length. Soliton propagation is employed where the controlling parameters lead to a balance between the pulse spreading due to dispersion and the pulse shrinking due to nonlinearity. The balance between the non-linearity effects from one side and the dispersion effects from the other side creates a solitary wave [24]. The dispersion of a medium (in the absence of non-linearity) makes the

various frequency components propagate at different velocities; while the non-linearity (in the absence of dispersion) causes the pulse energy to be continually injected, via harmonic generation, into higher frequency modes. That is to say, the dispersion effect results in broadening the pulse while the non-linearity tends to sharpen it. Analysis given in references [25, 26], the soliton bit rate per channel (B_{rs}) is given by:

$$B_{rs}^{-2} = 59.7 P_{so}^{-1} \left(\frac{\lambda_s}{1.54}\right)\left(\frac{A_{eff}}{20}\right)\left(\frac{3.2x10^{-20}}{n_{nl}}\right)\left|D_t\right| x10^6 \quad (51)$$

Where: (τ_{min}) is the minimum pulse broadening in nsec. The spectral and thermal sensitivities are the guide of the measurement the relative variations of the outputs and the relative variations of the inputs. The soliton and MTDM bit rate within thermal, loss, dispersion sensitivity coefficients, and parameters as (S_T^{Brs}), (S_α^{Brs}), (S_D^{Brs}), (S_T^{Brm}), (S_α^{Brm}), and (S_D^{Brm}) are taken into account as criteria of a complete comparison between silica doped and plastic materials and are given:

$$S_T^{B_{rs}} = \frac{T}{B_{rs}} \cdot \frac{dB_{rs}}{dT} = \frac{T}{B_{rs}} \cdot \frac{A_1^2 B_1 \left(3\lambda^3 B_1^2 + D_t \lambda_1^2\right)}{\left(\lambda^3 - A_1^2 B_2^2 + D_t^2 \lambda^2 A_2 B_2^3\right)} \quad (53)$$

$$S_\alpha^{B_{rs}} = \frac{\alpha}{B_{rs}} \cdot \frac{dB_{rs}}{d\alpha} = \frac{\alpha}{B_{rs}} \cdot \frac{0.633 B_1^2 \alpha_t \left(2\lambda_s^2 A_1 B_3^2 T^2 + \lambda_s^4 A_2^3 B_1^4 T^3\right)^2}{\left(A_1^2 \lambda_s T^3 + \alpha_t^3 B_2^2 \lambda_s^3\right)^3} \quad (54)$$

$$S_D^{B_{rs}} = \frac{D}{B_{rs}} \cdot \frac{dB_{rs}}{dD} = \frac{D}{B_{rs}} \cdot \frac{14.7 \lambda_s^2 A_{eff} \left(A_1^2 B_1^2 T^3 \lambda_s^3 + A_2^3 B_1^3 \lambda_s^2\right)^3}{n_{nl} P_{so} \left(A_1 B_2^3 \lambda_s^2 - T^2 B_2^2 A_1^4 \lambda_s^4\right)^2} \quad (55)$$

$$S_T^{B_{rm}} = \frac{T}{B_{rm}} \cdot \frac{dB_{rm}}{dT} = \frac{T}{B_{rm}} \cdot \frac{B_1^2 A_2 \left(4\lambda^2 A_1^3 + D_t^2 \lambda_1 A_2^3 B_2^3\right)}{\left(\lambda^2 + B_2^2 A_3^2 - D_t^3 \lambda B_2^2 A_2^4\right)^2} \quad (56)$$

$$S_\alpha^{B_{rm}} = \frac{\alpha}{B_{rm}} \cdot \frac{dB_{rm}}{d\alpha} = \frac{\alpha}{B_{rm}} \cdot \frac{0.633 A_1^2 \alpha_t \left(2\lambda_s^2 B_1 A_3^2 T^5 + \lambda_s^4 B_2^3 A_1^4\right)^2}{\left(B_1^2 \lambda_s \alpha_t^5 + T^2 A_2^2 \lambda_s^3\right)^3} \quad (57)$$

Where: (n_{nl}) is the nonlinear Kerr coefficient in m^2/watt,

(A_{eff}) is the effective area in μm^2/Watt,

(P_{so}) is the optical signal power in Watt,

(λ_s) is the optical signal wavelength in μm,

and (D_t) is the total chromatic dispersion coefficient in nsec/nm.km. Ref. [27] derived the condition for MTDM where the bit rate B_{rm} is given by:

$$B_{rm} = \frac{0.25}{\tau_{min}} Gbit/\sec \quad (52)$$

$$S_D^{B_{rm}} = \frac{D}{B_{rm}} \cdot \frac{dB_{rm}}{dD} \cdot = \frac{D}{B_{rm}} \cdot \frac{14.7 \lambda_s^2 A_{eff} \left(B_1^2 A_1^2 \lambda_s^3 + T^4 B_2^3 A_1^3 \lambda_s^2\right)^2}{n_{nl} P_{so} \left(T^2 B_1 A_2^3 \lambda_s^2 - A_2^2 B_1^4 \lambda_s^4\right)^3} \quad (58)$$

Using the series of the set of equations analysis from Eq. (1) to Eq. (59), the transmission bit rates per channel for both silica-doped and plastic materials are investigated under wide optical ranges which give the minimum values of both total losses and total dispersion. The calculations are based on the values:

Central optical signal wavelength (λ_0) is selected to be (1.2 μm ≤ λ_0 ≤ 1.65μm), relative refractive index difference for silica-doped material (Δn) is chosen to be ($0.006 \le \Delta n \le 0.008$), source spectral line width (Δλ) is selected to be (0.1 nm ≤ Δλ ≤ 0.5 nm), GeO_2 mole fraction (x) is chosen to be ($0.0 \le x \le 0.2$), relative refractive index difference for plastic material (Δn) is determined where ($0.05 \le \Delta n \le 0.07$), and fiber temperature (T) is chosen to be (290°K ≤ T ≤ 330°K).

4. Results and Discussions

We have analyzed the propagation problems of these materials in the interval of 1.2 μm to 1.65 μm under the set of affecting parameters at temperature range varies from 290 °K to 330 °K. The following numerical sets of data are used to obtain transmission bit rate and capacity-distance product per channel as follows: (1.2 μm ≤ λ_0 ≤ 1.65 μm); (λ_0): central optical signal wavelength, (0.1 nm ≤ Δλ ≤ 0.5 nm); (Δλ): spectral line width of the optical source, effective area; A_{eff}= 85 μm^2, (2 Km ≤ L ≤ 10 Km); (L): transmission distance, (0.0 ≤ x ≤ 0.2); (x): percentage of Germania doped in silica fibers, (0.05 ≤ Δn_{PMMA} ≤ 0.07); (Δn_{PMMA}): relative refractive-index difference, (0.006 ≤ $\Delta n_{silica\text{-}doped}$ ≤ 0.008), ($\Delta n_{silica\text{-}doped}$): relative refractive-index difference for silica-doped, (4 mwatt ≤ P_s ≤ 30 mwatt); (Ps): optical signal power. At the set of affecting parameters {optical signal wavelength (λ_s), ambient temperature (T_0), and relative refractive-index difference (Δn)}, both the effective performance of plastic, and Germania-doped silica fibers are processed based on the transmission bit rate or capacity-distance product per channel: Transmitted bit-rate **x** transmission distance (L) is given by:

$$P_r = B_r \times L \text{, Gbit.km/sec} \quad (59)$$

The transmitted bit-rate per optical channel is also a special criterion for comparison for different fiber cable materials of plastic and silica-doped fibers. Based on the clarified variations in figures (1- 20), the facts are assured:

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

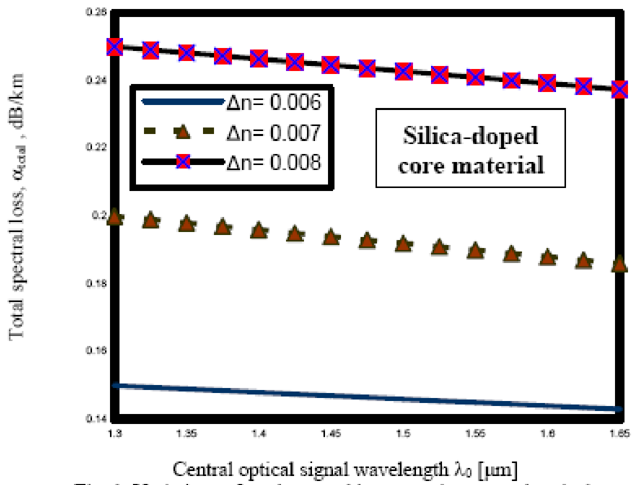


Fig. 1. Variations of total spectral losses against central optical signal wavelength at the assumed set of parameters.

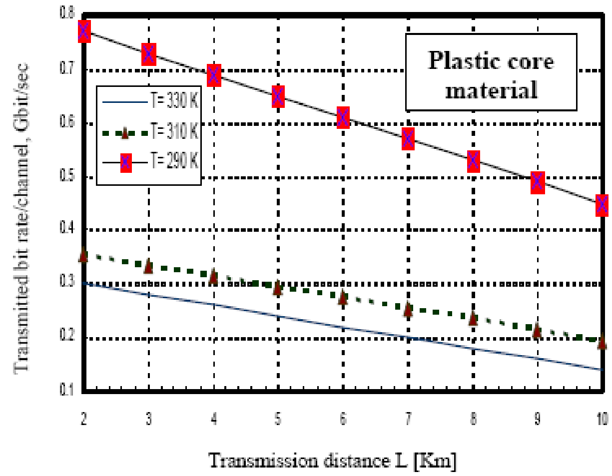


Fig. 4. Variations of transmitted bit rate per channel against variations of transmission distance at the assumed set of parameters.

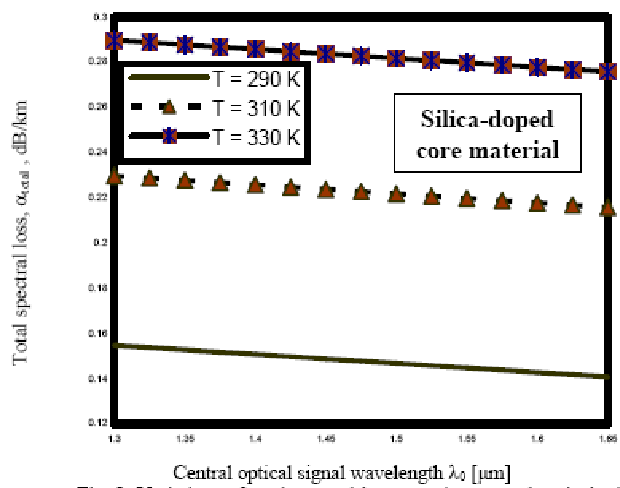


Fig. 2. Variations of total spectral losses against central optical signal wavelength at the assumed set of parameters.

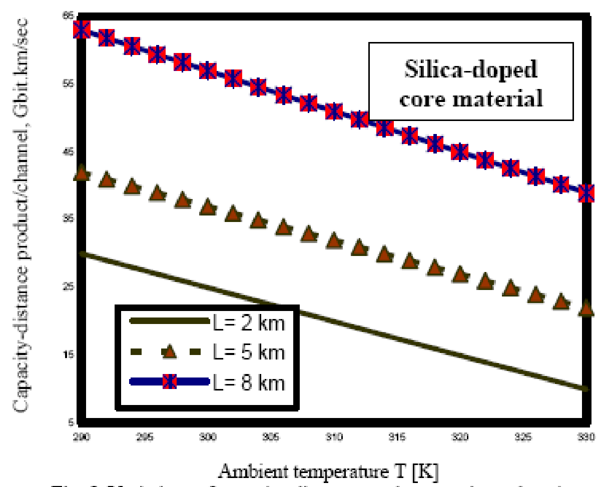


Fig. 5. Variations of capacity-distance product per channel against variations of ambient temperature at the assumed set of parameters.

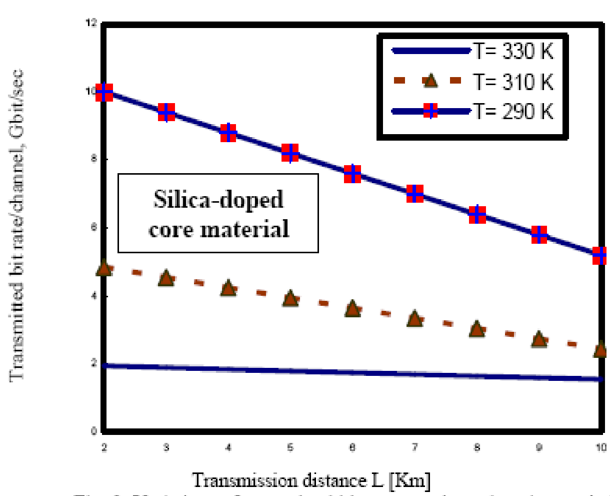


Fig. 3. Variations of transmitted bit rate per channel against variations of transmission distance at the assumed set of parameters.

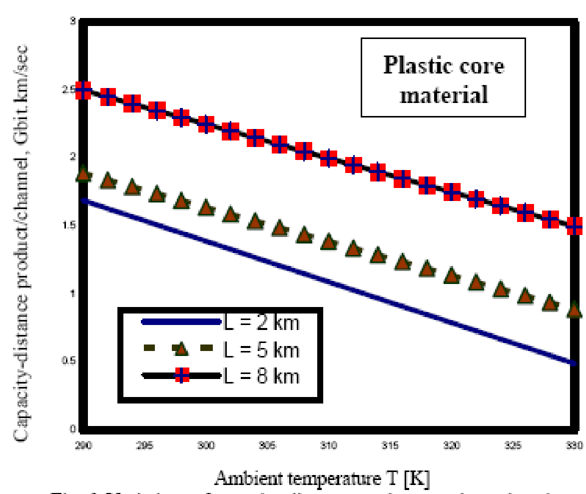


Fig. 6. Variations of capacity-distance product per channel against variations of ambient temperature at the assumed set of parameters.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

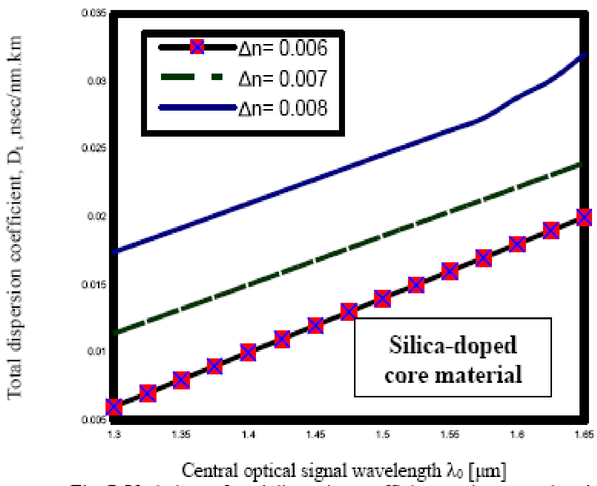


Fig. 7. Variations of total dispersion coefficient against central optical signal wavelength at the assumed set of parameters.

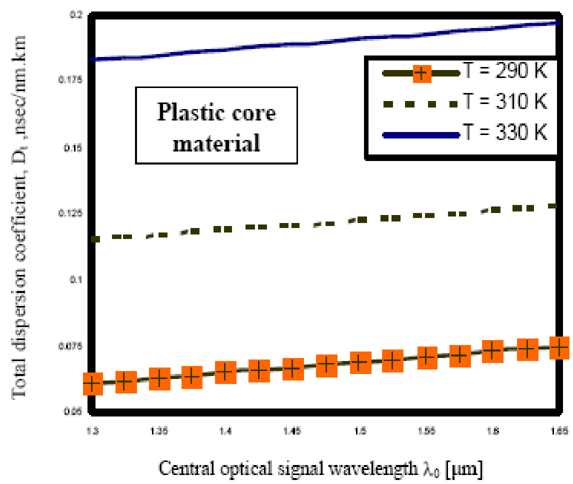


Fig. 10. Variations of total dispersion coefficient against central optical signal wavelength at the assumed set of parameters.

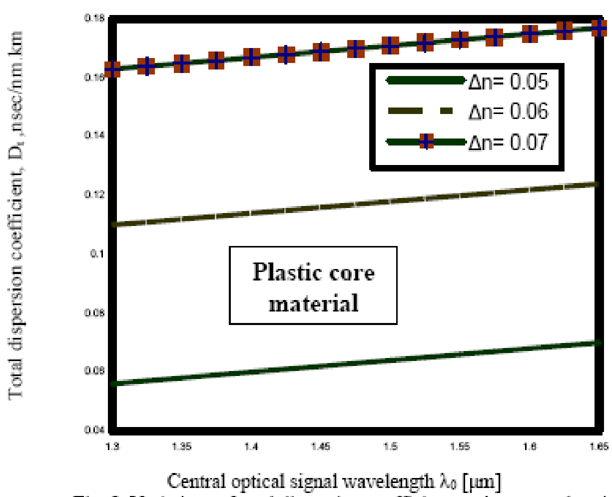


Fig. 8. Variations of total dispersion coefficient against central optical signal wavelength at assumed set of parameters.

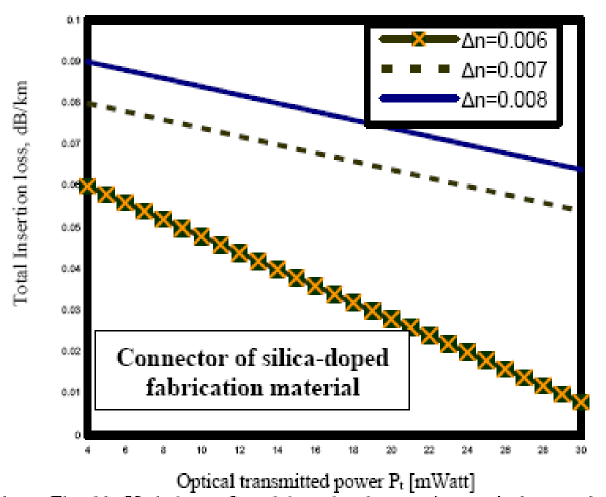


Fig. 11. Variations of total insertion loss against optical transmitted power at the assumed set of parameters.

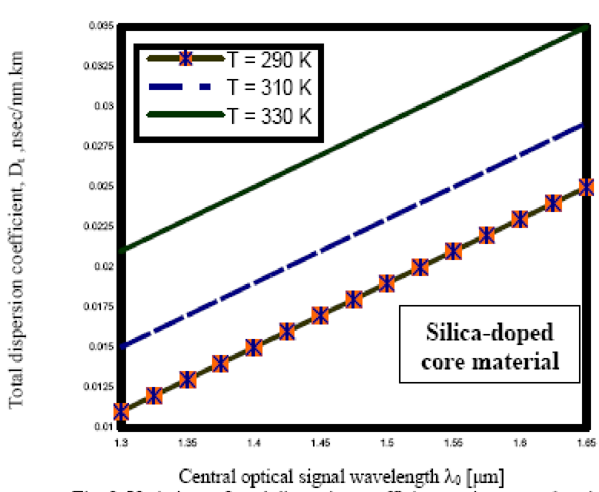


Fig. 9. Variations of total dispersion coefficient against central optical signal wavelength at the assumed set of parameters.

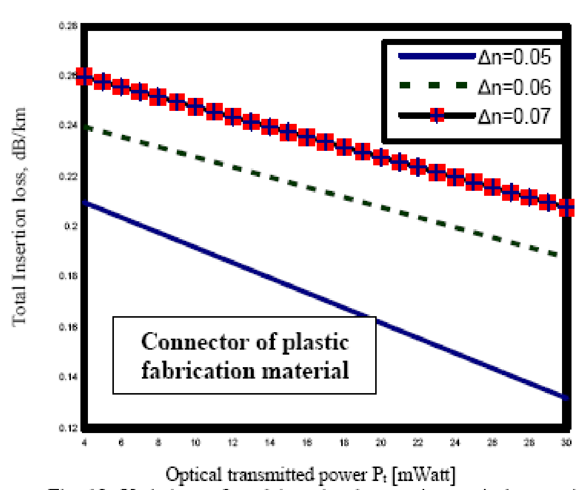


Fig. 12. Variations of total insertion loss against optical transmitted power at the assumed set of parameters.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

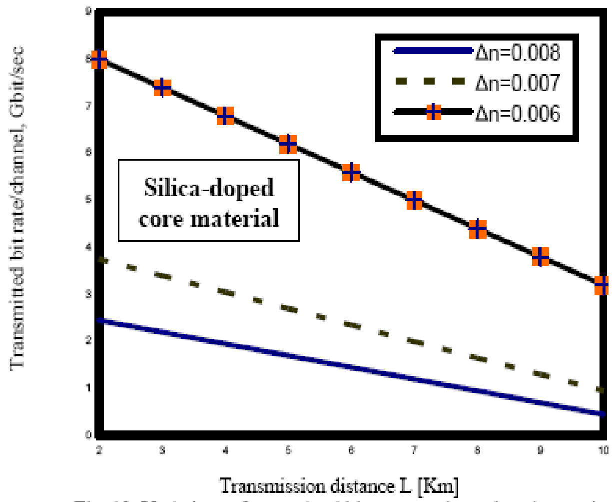


Fig. 13. Variations of transmitted bit rate per channel against variations of transmission distance at the assumed set of parameters.

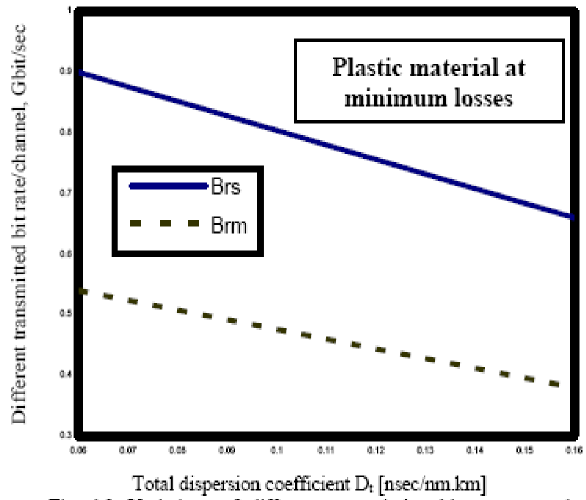


Fig. 16. Variations of different transmission bit rates per channel against total dispersion coefficient at the assumed set of parameters.

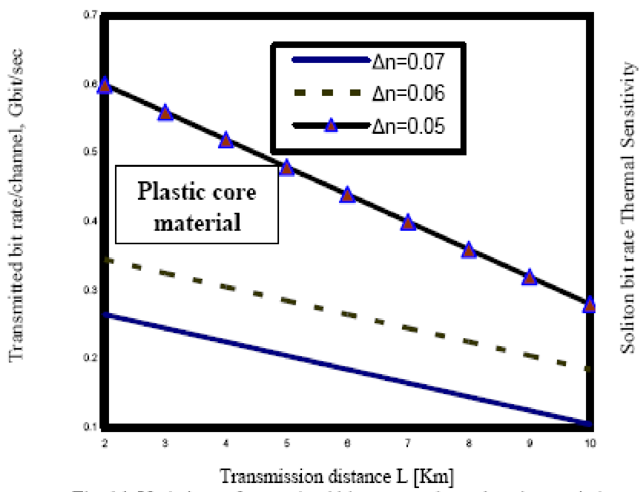


Fig. 14. Variations of transmitted bit rate per channel against variations of transmission distance at the assumed set of parameters.

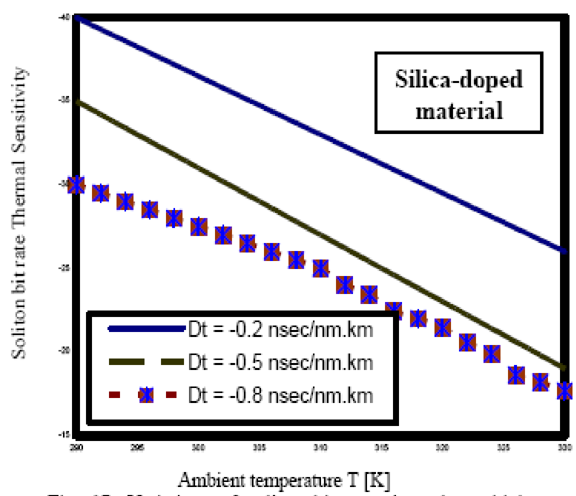


Fig. 17. Variations of soliton bit rate thermal sensitivity against ambient temperature at the assumed set of parameters.

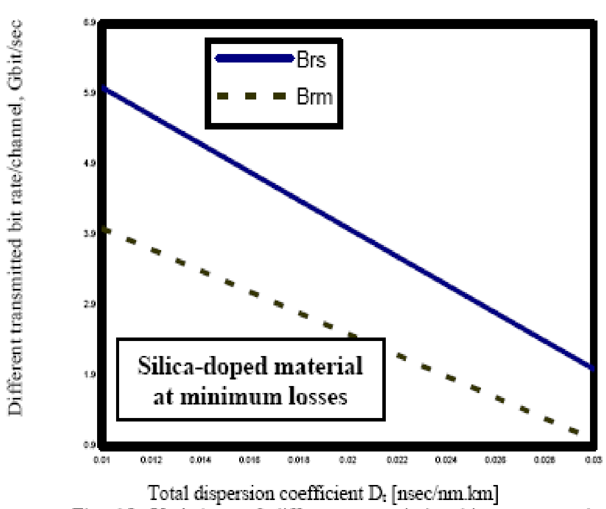


Fig. 15. Variations of different transmission bit rates per channel against total dispersion coefficient at the assumed set of parameters.

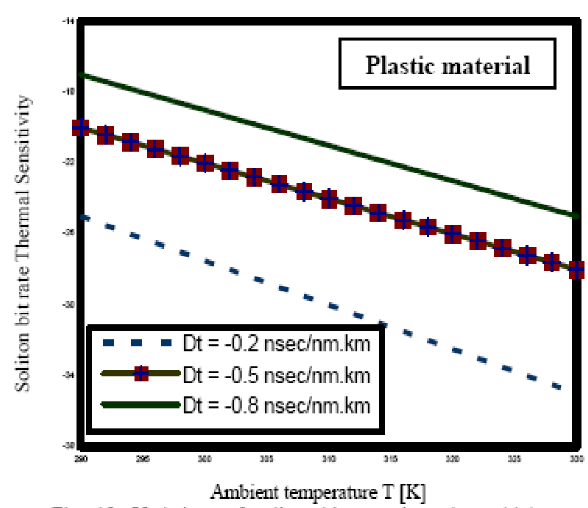


Fig. 18. Variations of soliton bit rate thermal sensitivity against ambient temperature at the assumed set of parameters.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

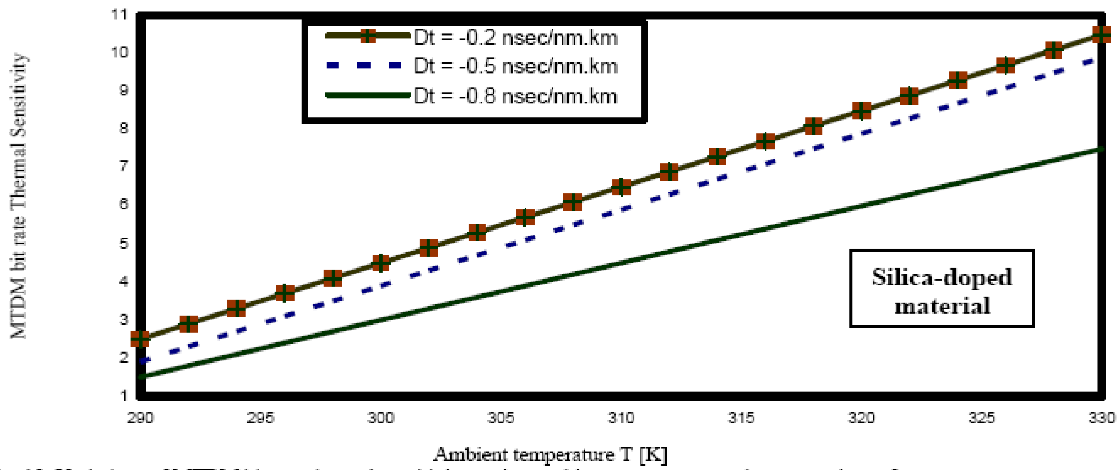


Fig. 19. Variations of MTDM bit rate thermal sensitivity against ambient temperature at the assumed set of parameters.

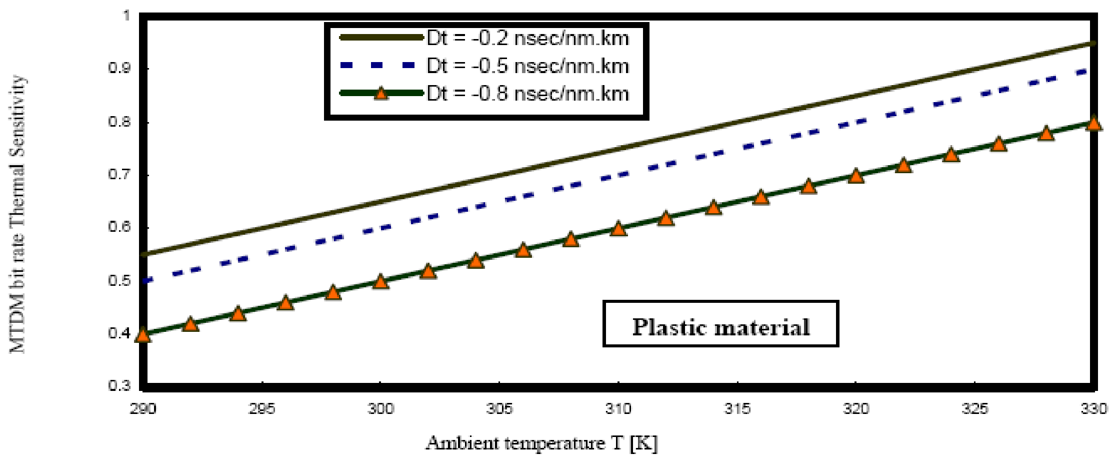


Fig. 20. Variations of MTDM bit rate thermal sensitivity against ambient temperature at the assumed set of parameters.

i- Figs. (1, 2) prove that as λ_0 increases, the total spectral losses decrease at constant both T, and Δn for silica-doped fibers. But as Δn, and T increase, the total spectral losses also increase at constant λ_0.

ii- Figs. (3-4) indicate that as L increases, the transmission bit rate per channel decreases for both silica-doped and plastic fibers at constant T. As well as T increases, the transmission bit rate per channel decreases at the constant L.

iii- Figs. (5-6) assure that as T increases, the capacity-distance product per channel decreases for both silica-doped and plastic fibers at constant L. Moreover as L increases, the capacity-distance product per channel also increases at constant both T.

iv- Figs. (7, 8) demonstrate that as λ_0 increases, total dispersion coefficient also increases for both silica-doped and plastic fibers at the constant Δn. As well as Δn increases, total dispersion coefficient also increases at constant λ_0.

v- Figs. (9, 10) prove that as λ_0 increases, total dispersion coefficient also increases for both silica-doped and plastic fibers at the constant T. As well as T increases, total dispersion coefficient also increases at constant λ_0.

vi- Figs. (11, 12) indicate that as the P_t increases, the total insertion loss decreases for both silica-doped and plastic fibers at constant Δn. As well as Δn increases, the total insertion loss also increases at constant P_t.

vii- Figs. (13, 14) demonstrate that as L increases, the transmission bit rate per channel decreases for both silica-doped and plastic fibers at constant Δn. But as Δn increases, the transmission bit rate per channel decreases at constant L.

viii- Figs. (15, 16) assure that as D_t increases, the soliton and MTDM bit rates decrease for both silica-doped and plastic materials, but with silica-doped presents higher bit rates than plastic materials at minimum losses.

x- Figs. (17, 18) indicate that as T increases, Soliton bit rate thermal sensitivity also increases at constant D_t, but with increasing D_t, we have observed that soliton bit rate thermal sensitivity deceases at constant T for both silica-doped and plastic materials.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

xi- Figs. (19, 20) prove that as T increases, MTDM bit rate thermal sensitivity also increases at constant D_t, but with increasing D_t, we have observed that MTDM bit rate thermal sensitivity increases at constant T for both silica-doped and plastic materials.

Therefore we can summarize reduction of the propagation problems within soliton transmission technique and offers the best conditions for the performance of transmission bit rates and capacity-distance products of silica-doped and plastic materials for different transmission distances in advanced local area optical communication networks for suitable operating conditions as shown in Table 1.

Table 1: Transmission bit rates and capacity-distance product for both silica-doped and plastic core materials.

Transmission Distance	Fiber cable core materials			
	Silica-doped material		Plastic material	
	Best conditions of operation at Δn = 0.006 and T = 290 ºK		Best conditions of operation at Δn = 0.05 and T = 290 ºK	
	Transmission bit rate/channel (Gbit/sec)	Capacity-distance product/channel (Gbit.km/sec)	Transmission bit rate/channel (Gbit/sec)	Capacity-distance product/channel (Gbit.km/sec)
L = 2 Km	8 Gbit/sec	30 Gbit.km/sec	0.5 Gbit/sec	1.5 Gbit.km/sec
L = 5 Km	5 Gbit/sec	45 Gbit.km/sec	0.3 Gbit/sec	2 Gbit.km/sec
L = 8 Km	2 Gbit/sec	60 Gbit.km/sec	0.1 Gbit/sec	2.5 Gbit.km/sec

5. Conclusions

The characteristics of both of silica doped and plastic fibers are investigated under the different affecting parameters. soliton and MTDM high transmission techniques are employed for reducing the propagation problems as limiting factors such as total losses and dispersion across silica-doped and plastic materials in (ALAOCN) within suitable affecting parameters. It is an evident that with the decreasing of both of an ambient temperature and relative refractive-index difference, it leads to higher transmission bit rate and capacity-distance product per channel across both of silica-doped and plastic materials. It is observed that the performance of transmission bit rates per channel of silica-doped materials will be higher than the plastic materials. When both of relative refractive-index difference and ambient temperature are constants with the increasing of the transmission distance then the capacity-distance product per channel will be increased in the silica-doped materials more than in the plastic materials. Moreover the increasing of the percentage of Germania doped in silica fibers leads to moving the zero dispersion at higher wavelength, λ=1.55 μm (minimum losses), then decreasing of the total dispersion and the total spectral losses and hence the transmission bit rates per optical channel in silica-doped fibers will be increased. Therefore we can say that the lowest total dispersion and total losses of silica-doped fibers make these fibers as the best candidate media for long haul optical transmission in advanced optical communication networks.

References

[1] M. Hossen, M. Asaduzzaman, and G. C. Sarkar, "Analysis of Dispersion of Single Mode Optical Fiber," National Conference on Communication and Information Security , pp. 143-146, NCCIS 2007.

[2] Abd El-Naser A. Mohammed, Abd El-Fattah A. Saad, and Ahmed Nabih Zaki Rashed, "Applications of Arrayed Waveguide Grating (AWG) in Passive Optical Networks," IJFGCN International Journal of Future Generation Communication and Networking, Vol. 2, No. 2, pp. 25-36, June 2009.

[3] M. Kovacevica, and A. Djordjevichb, "Temperature Dependence Analysis of Mode Dispersion in Step-Index Polymer Optical Fibers," Vol. 16, No. 4, pp. 649-651, Proceedings of the International School and Conference on Photonics 2009.

[4] L. N. Binh, L. H. Binh, and V. T. Tu, "Routing and Wavelength Assignment and Survivability of Optical Channels in Ultra-high Speed IP over DWDM Networks Under Constraints of Residual Dispersion and Nonlinear Effects," IJCSNS

International Journal of Computer Science and Network Security, Vol. 9, No. 2, pp. 49-60, Feb. 2009.

[5] S. A. Ahsan Rajon, and Md. Rafiqul Islam, "An Enhanced Energy-Efficient Data Communication Scheme for Resource Constrained Wireless Sensor Networks," International Journal of Computational Intelligence and Information Security, Vol. 1 No. 1, pp. 4-14, Jan. 2010.

[6] A. Ghatak and K. Thyagarajan, " An Introduction to Fiber Optics." Cambridge Univ. Press, 2008.

[7] H. Y. Choi, Paul K. J. Park, and Y. C. Chung," Chromatic Dispersion Monitoring Technique Using Pilot Tone Carried by Broadband Light Source," IEEE Photonics Technology Letters, Vol. 21, No. 9, pp. 578-580, May 2009.

[8] A. Sangeetha , S. K. Sudheer, and K. Anusudha, "Performance Analysis of NRZ, RZ, and Chirped RZ Transmission Formats in Dispersion Managed 10 Gbps Long Haul WDM Light Wave Systems," International Journal of Recent Trends in Engineering, Vol. 1, No. 4, pp. 103-105, May 2009.

[9] O. L. Ladouceur, H. Wang, A. S. Garg, and K. Bergman, "Low Power, Transparent Optical Network Interface for High Bandwidth Off Chip Interconnects," Optics Express, Vol. 17, No. 8, pp. 6550-6561, April 2009.

[10] A. S. Samra, and H. A. M. Harb, "Multi-layer Fiber for Dispersion Compensating And Wide Band Amplification," UbiCC Journal, Vol. 4, No. 3, pp. 807-812, August 2009.

[11] A. M. Rocha, B. Neto, M. Faveo, and P. S. Andre, "Low Cost Incoherent Pump Solution for Raman Fiber Amplifier," Optica Applicata, Vol. XXXIX, No. 2, pp. 287-293, 2009.

[12] M. S. Ab-Rahman, H. Guna, M. H. Harun, S. D. Zan and K. Jumari, "Cost-Effective Fabrication of Self-Made 1x12 Polymer Optical Fiber-Based Optical Splitters for Automotive Application," American J. of Engineering and Applied Sciences, Vol. 2, No. 2, pp. 252-259, 2009.

[13] L. N. Binh and L. H. Binh, " Transport of Assigned Wavelength Channels Over Ultra-high Speed Ethernet All-Optical DWDM Networks Under Constraints of Fiber Chromatic and Polarization Mode Dispersion Effects," JCSNS International Journal of Computer Science and Network Security, VOL. 9 No. 8, pp. 27-37, August 2009.

[14] S. S. Kalker, "Rapid Modeling and Estimation of Total Spectral Losses in Optical Fibers," J. Lightwave Technol., Vol. 4, No. 8, pp. 1125-1131, August 1986.

[15] T. Kaino, "Absorption Losses of Low Loss Plastic Optical Fibers," Jpn. J. Appl. Phys., Vol. 24, pp.1661-1163, 1988.

[16] H. Murofushi, "Low Loss Perfluorinated POF," in Proc. Fifth International Conference on PLAstic Optical Fibers and Applications-POF'96, Paris (France), pp.17-23, 1996.

[17] David K. Cheng, Field and Wave Electromagnetics (2nd edition), Prentice Hall, 1989.

[18] B. Saleh and M. C. Teich, Fundamental of photonics, Wiley-Interscience, 1991.

[19] W. Fleming, "Dispersion in GeO_2-SiO_2 Glasses," Applied Optics, Vol. 23, No. 24, pp. 4486-4493, 1984.

[20] S.T. Cundiff, B.C. Callings, L. Bovine, and W. H.Knox, "Propagation of Lightly Chirped Pulses in the Graded-Index Polymer Optical Fiber Systm, Toward Gigabit Data Links," Appl. Opt., Vol.13, No. 35, pp. 2048-2053, 1996.

[21] A. Tagaya, S. Teramoto, E. Nihei, K. Sasaki, and Y. Koike," High-Power and High-Gain Organic Dye-Doped Polymer Optical Fiber Amplifiers: Novel Techniques For Preparation and Spectral Investigation," Appl. Opt., Vol. 36, No. 28, pp. 572-578, 1997.

[22] T. Ishigure, E. Nihei, and Y. Koike, "Optimum Refractive Index Profile of the Graded-Index Polymer Optical Fiber, Toward Gigabit Data Links," Appl. Opt., Vol. 35, No.12, pp. 2048-2053, 1996.

[23] J. Senior, Manchester Polytechnic, Optical Fiber Communications Principles and Practice–Hall International, Inc., London, 1985.

[24] R. Gangwar, S. P. Singh, and N. Singh, "Soliton Based Optical Communications," Progress In Electromagnetics Research (PIER), Vol. 74, pp. 157–166, 2007.

[25] T. Lakoba, and G. Agrawal, "Optimization of the Average Dispersion Range for Long–Haul Dispersion Managed Soliton Systems ," J. Lightwave Technol., Vol. 18, No. 11, pp. 1504-1512, 2000.

[26] T. Hoshida, T. Terahara, and H. Onaka "Specification Method for Loss Distributed Profile in Transmission Paths for WDM Transmission System Employing Distributed Raman amplification," Optical Network Mag., Vol. 2, No. 2, pp. 86–91, Sep./Oct., 2001.

[27] T. Otani, T. Miyazaki, S. H.. Carassa, and S. Yamamot, "40-Gb/sec Optical 3R Regenerator Using Electro Absorption Modulators for Optical Communication Networks, " J. Lightwave Technol., Vol. 20, No. 2, pp. 195-200, 2002.

Classification Maintenance Requests in Bug Tracking System

Naghmeh Mahmoodian
University Putra Malaysia
Faculty of computer science and information technology, UPM, 43400 upm serdang, selongor Malaysia
Kuala Lumpur, Malaysia
naghmeh.ma@gmail.com

Rusli Abdullah
University Putra Malaysia
Faculty of computer science and information technology, UPM, 43400 upm serdang, selongor Malaysia
Kuala Lumpur, Malaysia
rusli@fsktm.upm.edu.my

Masrah Azrifah Azim Murad
University Putra Malaysia
Faculty of computer science and information technology, UPM, 43400 upm serdang, selongor Malaysia
Kuala Lumpur, Malaysia
masrah@fsktm.upm.edu.my

***Abstract*—Complex process of software modification s called software maintenance (SM), is a costly and time consuming task. Classification of maintenance request (MR) on the type MR which are corrective, adaptive, perfective or preventive. The maintenance type (MT) are important in keeping the quality factors of the system. However, classification of MT is difficult in nature and this affect maintainability of the system. Thus, there is a need for tool which is able to classify MRs into MT automatically. In this study, we present new result of combination texts of features during MR classification. Three different classifications techniques, namely Decision Tree, Naïve Bayesian and Logistic Regression are applied to perform classification. 600 MRs from Mozilla bug tracking system (BTS) are used as source of data. The results show that the model is able to classify MRs into MT with accuracy between 67.33 and 78.17%.**

Keywords— Classification; Software Maintenance; Maintenance Type; Classification; Corrective Maintenance

I. INTRODUCTION

In the software maintenance (SM) area when the problem occurs the issues in SM are being managed by computer software called the BTS.

BTS is a database of reported errors, commonly referred as bugs. BTS contains information such as title, description, reporter, source of request, time and date of error.

MR is reporting error in BTS, which may be classified into MT. MT, is useful in determining quality factors of a software system.

Software maintenance, as defined by IEEE is as follow [14]:

“Modification of a software product after delivery to correct faults, improves performance or other attributes, or adapts the product to a modified environment.”

Previous author also mentioned to the software maintenance definition as [1]:

The whole group of activities that are required to support a software system in a cost-effective manner.

SM is the key to product quality and risk reduction. It guarantees reliability, capability, availability, efficiency and safety [2].SM is a time consuming and costly. In some cases, the cost of maintaining application software is higher than the original development cost.In practice, it is difficult to classify maintenance activities due to independent classification of SM activities [4].

However, traditional text-based classifier are usually based on manual categorization, which is very time consuming and is difficult to maintain.

Automatic classification methods are very much needed for assigning bugs to developer, building fault-prone model, classifying MR, and scheduling SM activities such as enhancement or restructuring.

Machine learning (ML) techniques with powerful features are well-suitable for SM process in a textual environment. The techniques may be used to predict or classify SM fault and to classify MR into different forms such as number of fault and type of fault.

Textual information in BTS is used to identify

corrective maintenance from other types of SM, which are adaptive, perfective, and preventive. Thus, it is worth to classify MRs into MT without any human interaction in order to reduce the time and cost of SM task.

This research presents the result of automatic classification system that could classify MRs into "corrective" and "no-corrective". Bayesian classifiers and Decision Tree model and Logistic Regression were used to make the classifiers model. In this area examined the text of 600 MRS which is reported to Mozilla BTS. We labeled the MRs into two classes by using four types of information for each MR which is reported. It is useful to determining type of the maintenance and it is also able to improving software quality. MR classification is much close to effect on determining the corrective, adaptive, perfective and preventive of software maintenance. Type of software maintenance is one of the factors that is important for maintainability of the system without any damage on software architecture.

This research is accomplishing SM task automatically by extracting features or information from the texts such as title, description, reporter, error encountered. Finally MRs will be assignee into one of two existent categories.

Thus, we support maintainer in their daily activities with classifying MRs into MT automatically and correctly. The framework of classifying MR into maintenance type [19] is used as a source for classification. Machine learning techniques to classify requests of Mozilla is utilized to achieve a high percentage of correctly classified requests for the two classes of issues that are referred to as **"**corrective**"** and **"**no-corrective**"** maintenance.

The reminder of the article is organized as follows: section II reviews related works that are conducted in software maintenance type and machine learning techniques, section III includes the preliminaries of the MT,BTS and WEKA tool. Section IV describe the propose model to classify MR into MT. Section V explains the experimental results and discussion in particular show the result of the machine learning models namely as naïve Bayesian, decision tree and logistic regression on BTS. Finally, section VI outlines the conclusion and VII presents the future work.

II. RELATED WORK

There are three parts of the literature organized in this study as follow:

A. Software maintenance

In this section, a number of previous works in SM process are reviewed:

[6], describes a model of SM that investigates the risks and depicts the activities executed by measuring the impact of change request. Incoming texts represent the measurements with information that can be used for decision about when and how to make a change. This is where that importance of MT is determined.

Software maintenance is indeed a part of software engineering activities that requires more effort as compared to other activities in software engineering. The main goal of SM is to correct, to adapt or to enhance. Preventive maintenance is important to improve maintainability of the system. For this reason, researchers are trying to reduce SM effort by automating SM task [7].

[8], highlights importance of change in software and how to characterize the effect of change on software such as the cause of the change, the type of the change that needs to be made, and the part of the system that requires change. However, no suggestions are put forward to recognize varieties of MTs. Often, changes in the system have chain effects and will cause the next change in the system. Hence, managing change is expensive due to great effort required.

B. Type of software maintenance

Software maintenance is divided into various categories by different researchers in this area.

[9], expresses the base of SM activities, which is divided into three types: corrective, adaptive, and perfective. Corrective maintenance is performed in response assessment failure, which contains of performance failure, implementation failure, and processing failure. Adaptive maintenance is change in data and processing environment, which contains of change in data environment and change in processing environment. Perfective maintenance aim is to make the program into a more perfect design implementation and to enhance performance while improve maintainability.

[10], proposes maturity model for daily SM activities to improve SM functions. They first illustrate SM process and classify SM activities based on ISO/IEC 12207. This model contains the activities and tasks of the maintainer when the system goes into modification due to errors or problems. The improved maintenance model facilitates customer

satisfaction through daily maintenance activities provided. However, no comparisons were made between the new model and the existing standard model for SM quality evaluation.

Nonetheless, categorizing change requests into MT is not an easy task for software maintainer [11]. Most often, the software maintainer fails to initialize the correct category of change request into MT, while it is important to do the task correctly. While, change request is used to estimate MT and duration of maintenance, MR is used in our paper.

C. *Classification of Software Maintenance*

There are various models and techniques that are used in different research to classify software problems.

Evaluation of change massage classification is performed by using Kappa coefficient [12], which is a technique to measure the agreement between manual and automatic classification. Consequently, Stuart-Maxwell test [13] is used due to different tendencies to classification category. As the result, [15] manages to save 70% of the maintainer time through automatic classification of change messages. The maximum accuracy obtained from the experiments was only 63%.

ML aid SM in increasing quality factors of software systems, for example reliability and maintainability. Most economic damage is mainly caused by software failures [16]. Automatic classification helps to classify requests and to direct them to the maintenance in a short time frame, hence improving the service rate. In this research, five ML techniques are used to perform classification task, which are Vector Space Model (VSM), Bayesian networks, SVM, K-nearest neighbor, and regression three. The research also incorporates intelligent agent to route the issues to specific maintenance team.

Statistic methods such as logistic and linear regression are also being used to analyze open source software in terms of the quality [17]. The logistic regression method is used to classify whether a class contains bugs or otherwise, while linear regression is used to predict the number of bugs in each class. Decision tree and neural network are also used in classification to predict fault-proneness of the software code, instead of using source code metrics. The codes are classified into a class "without bugs" and a class "with bugs".

Among existing research on text classification in SM, we follow [24]. Antoniol and his colleagues use three ML techniques, which are Decision Tree, naïve Bayesian, and logistic regression to classify the request texts into "bug" or "non-bug".

III. PRELIMINARIES

In the following sub-section, introduces the general concepts like phases in MT, BTS and WEKA tool.

A. *Maintenance Type (MT)*

ISO/IEC 14764 (2006) divides MT into four types:

Corrective: Reactive modification of a software product performed after delivery to correct discovered problems.

Adaptive: Modification of a software product performed after delivery to keep a software product usable in a changed or changing environment.

Perfective: Modification of a software product after delivery to sustain performance and maintainability.

Preventive: Modification of a software product after delivery to detect and correct latent faults in the software product before them become effective faults.

B. *Bug Tracking System (BTS)*

A bug tracking system is a software application that is used for tracking reports on system errors or bugs in a software environment. The system is mainly used by programmers and quality assurance teams. BTS also acts as a database that records all information in the form of text about bugs in the system as indicated by [23].

Reported bug time, the bug title, severity that indicates the impact of the failure, the person who report the error, who worked on this error, description of error, error encountered that point to the product which is affected by the bug [18].

C. *WEKA Tool:*

WEKA is a tool that is developed at the University of Waikato for performing ML in Java programming language as illustrated in Figure 2.2. WEKA is a tool with ready algorithms for data analyses, classification and prediction, as well as easy to navigate graphical user interface. WEKA has been used in the area of SM by a number of researchers [20].

IV. MR Classification into MT

Previous researchers provided tool on the MR which could be use for daily maintenance activities [3, and 5].

When fault occurs in a software system, MR is released to the maintenance team who performs analysis on the MR description [21]. Type of requests

should be determined by the maintainer based on their expertise and the text of MR.

Human expert is one of the important evaluations necessary in text classification. Therefore, the data in this research are classified by three human experts. The text of MR is categorized based on key words by human experience. The results of manually classification get done through simple majority vote and decision on the MR of open-source system. In our work, MRs should be automatically assigned to relate MT which is provided by experts and were contained with composes of features. Naïve Bayesian, decision tree and logistic regression from various techniques are chosen for computing accuracy of model. 600 request of the standard dataset namely Mozilla is selected to implement our model. Mozilla is an open-source project has been used to build internet application such as web browser (Firefox), mailers and newsreader. Mozilla is developed in C++.

Metric for evaluating the performance of MRMT system have been presented. In order to evaluate accurate and inaccurate categorization of MRs, MLTs information is required. Accuracy is one of the most common metric which is used in machine learning techniques. We examine 600 MRs Classification, which is released by users, for new features. WEKA tool is utilized to perform three MLTs on model to show the advantage of it in categorizing MRs into MT. Now we are going to explain in more details the process of research to solve problem statement.

Mozilla, BTS is online system that user entered their requests into these systems. As extracting manually online data is a time consuming task, VB.net program is used to extract the requests (issues) from Mozilla accumulate as excel and text file. For this purpose the requests number is entered in issues box and cod extracted the texts of features which are determined in each BTS such as title, description. Request in BTS are identified by the terms of features. Two approaches are used for this research: first step is searching for terms of features manually that is time consuming and need great efforts of experts in this area. The second step is automatically determining MT. Thus, we start from first step to achieve the second step that is faster and without any human interaction. Features searching is used as an approach such as "fix", "error", "new". Manually classification of MRs by experts is used to pre-labelling requests for labelled as “corrective” or “non-corrective”. For example, state "Error thrown when requested and default language content of a file don't exist" is labelled as corrective.

In this paper, MR is classified to effectively categorize requests as “corrective” or “non-corrective”. This is possible because the features extracted from the BTS are seen to have high accuracy in classifying MR. Thus, developer has the opportunity to easily, automatically and immediately define the MT.The MR classification involves two steps: training and classification. The classification algorithms learn from a training set, that is, a collection of requests that are known to belong to an existing class (which is identified by human expert at the first step), that is, the requests are labelled with the known class. Features are extracted from the BTS and the classification algorithm learns which features are the most useful for discriminating among the two classes. In this context, feature often means some property of MR which is contain in request, such as title of error, description, error encountered, and source of request. The frequencies of terms for each feature in document are used as index word or keywords. Although, removing very frequent words (stop words) are used by researchers in text classification but we did not apply it in our work (the details explain in Section 4.4.4). Thus, all terms exist in projects are used as text features in our work

We apply the Linguistic features such as text filtering, stemming and indexing. In filtering, remove punctuation and splitting email address and camel-case identifier, and in continue the aim of stemming is removing plural and identifying the infinitive of verb. At the last step for indexing used Tf-idf indexing to find the frequency of the word to distinguish corrective from non-corrective and finally each request of BTS is labelled by c={0,1} will be use by MLT. A value of zero assigned to indicates a decision to corrective, while a value of one indicates a decision not corrective. When two human experts decide whether to classify the request under category of corrective, they may disagree, and this fact is effective on automatic classifier and happens with relatively high frequency. Three machine learning are briefly explained in follow:

Naïve Bayesian:Based on [22] Bayesian classifier is a statistical classifier that can predict class membership probability. In Naive Bayesian classifier the effect of attributes on a given class is independent of the other attributes. The model computes the

probability that a request is related to class c.

Where there are two classes, c= (0,1) and each tuple is represented by n-dimensional attribute A= (a_1,a_2,....a_n). Thus, A belongs to the class having the highest probability, condition on A. The probability of A written as follow:

$$P(C|a1,\ldots,an) = \frac{p(\mathrm{a1},\ldots,\mathrm{an|ci})}{\sum_{h=1}^{m} p(\mathrm{a1},\ldots,\mathrm{an|ch})} p(ci)$$

For a given class of requests A which all attributes are conditionally independent, all incoming requests ci are ranked by the condition probabilities:

$$p(ck|a1,\ldots,an) = \frac{\prod_{j=1}^{n} p(aj|ck)}{\sum_{h=1}^{q} \prod_{j=1}^{n} p(aj|ch) p(ch)} p(ck).$$

Decision Tree: Decision tree is top-down approach which starts at the root of the tree with a training set of tuples X= (a1, ..., an) and goes through the internal nodes, and the corresponding edge are followed. At the end the last node which is leaf determining the label c of the X. The leaf nodes are labelled with either 0 or 1 that indicated to corrective or adaptive. In this research we applied the Alternating Decision tree or AD tree. AD tree "data structure" and "algorithm" are a generalization of decision tree and have connections to boosting which is based on the question.

Logistic Regression: Logistic regression is a useful way of describing the relationship between one or more independent variables and a binary response variable that has only two possible values. Thus, class assumes only two values either 0 or 1.

There are some text of information on the source of request and error encountered that are most benefit to classifying and categorize a distinction between different type of MR. Combining the source of request and error encountered with the textual features of title, description might result in better classifying and learners. The results show that features might be in correlated. We find that some sources of requests are more interesting in particular type of maintenance, and also the most of errors which occurred in the same place caused the same MT. Thus, the combination of new features is helpful to improve the accurate classification on MT. The dataset has run against combination of the features, and each of the machines learning is used to find the accuracy for each of them. The details on other features and evaluation conditions will be illustrated in more detail in next section..

V. THE EXPREMENTAL RESULT

There are some tools that we have been used. Tools are useful to automatically classifying MRs into MT, such as stemming, Tf-idf, and WEKA Tool. Before we can use WEKA for classification, we need to prepare of data in WEKA format with ARFF extension WEKA tool was used for implementing Bayesian classifier, Decision Tree, and Logistic regression. The standard algorithms alternating naive Bayesian classifier, alternating decision tree and logistic regression was used to automatic classification on a set of BTS issues. Finally, cross validation was used to evaluate the performance. Especially, Ten-fold cross validation was used to select the subset of training set since this affects classification accuracy. The stemming was implemented by Porter stemmer from lsa package of the statistical environment (http://www.r-project.org). The Tf-idf indexing was used to find the frequency of the words that guide the classification technique to distinguish corrective.

This part is describing the result of the combination of new features, totally four features. We examined 20 and 50 selected text of features on Mozilla open source project. We show the correct classification on different number of 20and 50features selection as well as precision and recall for both classes.

In Table I, the naïve Bayesian classification when increasing the number of features the number of precision and recall is increased. Increasing number of features also can increase the accuracy of the model from 67.33 to 73.67.

TABLE I. Naïve Bayesian Classification

Select feature		**Naïve Bayes**		
20		***Predict***		
		Corrective	***Non-Corrective***	Rec
	Corrective	209	124	63%
	Non-corrective	72	195	73%
	Prec	74%	61%	**67.33%**
50		***Predict***		
		Corrective	***Non-Corrective***	Rec
	Corrective	236	97	71%
	Non-corrective	61	206	78%
	prec	80%	68%	**73.67%**

In Table II, decision tree classification with 20 number of feature it cannot distinguish corrective type but with increasing number of feature we can obtain better result and improving the accuracy from

68.5% to 68.7%. In this model with increasing the number of feature precision and recall decreased for corrective type.

If we comparing the naïve Bayesian and ADtree we can see the accuracy of the model with maximum number of feature is 73.67% and 68.7% that show the better result in naïve Bayesian model in Table I.

TABLEL II. Decision Tree Classification

Select feature		ADtree		
20		***Predict***		
		Corrective	***Non-Corrective***	Rec
	Corrective	258	75	78%
	Non-corrective	114	153	57%
	Prec	69%	67%	**68.5%**
50		***Predict***		
		Corrective	***Non-Corrective***	Rec
	Corrective	198	135	60%
	Non-corrective	53	214	80%
	prec	79%	61%	**68.7%**

In Table III, Logistic Regression classification with 20 number of feature it can distinguish corrective type but with increasing number of feature we can obtain better result and improving the accuracy from 67.5% to 78.17%. In this model with increasing the number of feature precision and recall decreased for corrective type.

In comparison with other model this model has better result on 50 texts of features.

TABLEL III: Logistic Regression Classification

Select feature		Logistic Regression		
20		***Predict***		
		Corrective	***Non-Corrective***	Rec
	Corrective	203	130	61%
	Non-corrective	65	202	76%
	Prec	76%	61%	**67.5%**
50		***Predict***		
		Corrective	***Non-Corrective***	Rec
	Corrective	79	115	41%
	Non-corrective	16	390	96%
	prec	83%	77%	**78.17%**

VI. CONCLUSION

Using the naïve Bayesian, decision tree and logistic regression to classify the system applications was developed in this work.

We automated classifying MRs into maintenance type. In the Bayesian classification, we applied different numbers of features and its effect on the accuracy of the model that indicated in the previous section. Increasing the numbers of features in the system could accurately classify the requests. We compared three machine learning methods and the result show that the Logistic Regression model has a better result in comparison with the other methods. We showed, combination of textual filed of MRs could be enough to classify them in to corrective and non-corrective.

As a future work, considering the semantic of the requests in classification can be directed a wider range of future research. On the other, feature can be used but need a new dictionary for classifying requests, while accuracy can be used as a metric for evaluation accurate classification of the model.

REFERENCES

[1] T.M. Pigoski, Practical software maintenance: best practices for managing your software investment, John Wiley & Sons, Inc. New York, NY, USA, 1996.

[2] M. Pariazar, J. Shahrabi, M.S. Zaeri, and Sh. Parhizi, "A Combined Approach for Maintenance Strategy Selection," 2008.

[3] J.F. Bowring, J.M. Rehg, and M.J. Harrold, "Active learning for automatic classification of software behavior," SIGSOFT Softw. Eng. Notes, vol. 29, 2004, pp. 195-205.

[4] E. Burch and Hsiang-Jui Kungs, "Modeling software maintenance requests: a case study," Software Maintenance, 1997. Proceedings., International Conference on, 1997, pp. 40-47.

[5] G.A. Di Lucca_, M. Di Penta__, S. Gradara, 2002, An Approach to Classify Software Maintenance Requests, Proceedings of the International Conference on Software Maintenance (ICSM.02), 2002 IEE

[6] S.L. Pfleeger, *Software Engineering: Theory and Practice*, Prentice Hall, 2001.

[7] Y. Singh and B. Goel, "A step towards software preventive maintenance," SIGSOFT Softw. Eng. Notes, vol. 32, 2007, p. 10.

[8] B.J. Williams and J.C. Carver, *Characterizing Changes to Assess Architectural Impact*, Citeseer, .

[9] E.B. Swanson, "The dimensions of maintenance," *Proceedings of the 2nd international conference on Software engineering*, San Francisco, California, United States: IEEE Computer Society Press, 1976, pp. 492-497.

[10] A. April, J.H. Hayes, A. Abran, and R. Dumke,

“Software Maintenance Maturity Model (SM^{<FONT SIZE='-1'>mm</FONT>}): the software maintenance process model,” *Journal of Software Maintenance and Evolution: Research and Practice*, vol. 17, 2005, pp. 197-223.

[1] L. Hatton, “How Accurately Do Engineers Predict Software Maintenance Tasks?,” *Computer*, vol. 40, 2007, pp. 64-69.

[12] W.D. Thompson and S.D. Walter, “A rEAPPRAISAL OF THE KAPPA COEFFICIENT,” *Journal of Clinical Epidemiology*, vol. 41, 1988, pp. 949-958.

[13] B. Everitt, *The analysis of contingency tables*, CRC Press, 1992.

[14] IEEE Standard for Software Maintenance. IEEE Std 1219-1993.

[15] A.E. Hassan, “Automated classification of change messages in open source projects,” *Proceedings of the 2008 ACM symposium on Applied computing*, Fortaleza, Ceara, Brazil: ACM, 2008, pp. 837-841.

[16] G.D. Lucca, “An Approach to Classify Software Maintenance Requests,” *Proceedings of the International Conference on Software Maintenance (ICSM'02)*, IEEE Computer Society, 2002, p. 93.

[17] T. Gyimothy, R. Ferenc, and I. Siket, “Empirical Validation of Object-Oriented Metrics on Open Source Software for Fault Prediction,” *IEEE Trans. Softw. Eng.*, vol. 31, 2005, pp. 897-910.

[18] M. Fischer, M. Pinzger, and H. Gall, “Populating a Release History Database from Version Control and Bug Tracking Systems,” Proceedings of the International Conference on Software Maintenance, IEEE Computer Society, 2003, p. 23.

[19] N. Mahmoodian, R. Abdullah, M.A. Azmi-Murad , “A Framework of Classifying Maintenance Requests Based on Learning Techniques”

[20] A. Hindle, D. German, M. Godfrey, and R. Holt, “Automatic classification of large changes into maintenance categories,” *2009 IEEE 17th International Conference on Program Comprehension*, Vancouver, BC, Canada: IEEE, 2009, pp. 39, 30.

[21] G.D. Lucca, “An Approach to Classify Software Maintenance Requests,” Proceedings of the International Conference on Software Maintenance (ICSM'02), IEEE Computer Society, 2002, p. 93.

[22] G.D. Lucca, “An Approach to Classify Software Maintenance Requests,” Proceedings of the International Conference on Software Maintenance (ICSM'02), IEEE Computer Society, 2002, p. 93.

[23] M. Kajko-Mattsson, “Common Concept Apparatus within Corrective Software Maintenance,” Proceedings of the IEEE International Conference on Software Maintenance, IEEE Computer Society, 1999, p. 287.

[24] G. Antoniol, K. Ayari, M.D. Penta, F. Khomh, and Y. Guéhéneuc, Is it a bug or an enhancement?: a text-based approach to classify change requests, Proceedings of the 2008 conference of the center for advanced studies on collaborative research: meeting of minds, Ontario, Canada: ACM, 2008, pp. 304-318.

(IJCSIS) International Journal of Computer Science and Information Security,
Vol. 8, No. 4, July 2010

An Optimized Clustering Algorithm Using Genetic Algorithm and Rough set Theory based on Kohonen self organizing map

[1]Asgarali Bouyer, [2]Abdolreza Hatamlou
[1,2]Department of Computer Science
[1]Islamic Azad University – Miyandoab Branch
[1]Miyandoab, Iran
[2]University Kebangsaan Malaysia
[2]Selangor, Malaysia
[1]basgarali2@live.utm.my, [2]hatamlou@iaukhoy.ac.ir

[3]Abdul Hanan Abdullah
Department of Computer and Information Systems,
Faculty of Computer Science and Information Systems,
[3]Universiti Teknologi Malaysia
81310 Skudai, Johor Bahru, Malaysia
[3]hanan@utm.my

***Abstract*—The Kohonen self organizing map is an efficient tool in exploratory phase of data mining and pattern recognition. The SOM is a popular tool that maps high dimensional space into a small number of dimensions by placing similar elements close together, forming clusters. Recently, most of the researchers found that to take the uncertainty concerned in cluster analysis, using the crisp boundaries in some clustering operations is not necessary. In this paper, an optimized two-level clustering algorithm based on SOM which employs the rough set theory and genetic algorithm is proposed to defeat the uncertainty problem. The evaluation of proposed algorithm on our gathered poultry diseases data and Iris data expresses more accurate compared with the crisp clustering methods and reduces the errors.**

Index Terms- SOM, Clustering, Rough set theory, Genetic Algorithm.

I. INTRODUCTION

The self organizing map (SOM) proposed by Kohonen [1], has been widely used in industrial applications such as pattern recognition, biological modeling, data compression, signal processing and data mining [2]-[5]. It is an unsupervised and nonparametric neural network approach. The success of the SOM algorithm lies in its simplicity that makes it easy to understand, simulate and be used in many applications. The basic SOM consists of neurons usually arranged in a two-dimensional structure such that there are neighborhood relations among the neurons. After completion of training, each neuron is attached to a feature vector of the same dimension as input space. By assigning each input vector to the neuron with nearest feature vectors, the SOM is able to divide the input space into regions (clusters) with common nearest feature vectors. This process can be considered as performing vector quantization (VQ) [6]. Also, because of the neighborhood relation contributed by the inter-connections among neurons, the SOM exhibits another important property of topology preservation.

Clustering algorithms attempt to organize unlabeled input vectors into clusters such that points within the cluster are more similar to each other than vectors belonging to different clusters [7]. The clustering methods are of five types: hierarchical clustering, partitioning clustering, density-based clustering, grid-based clustering and model-based clustering [8]. The rough set theory employs two upper and lower thresholds in the clustering process which result in a rough clusters appearance. This technique also could be defined in incremental order i.e. the number of clusters is not predefined by users.

Our goal is to optimized clustering algorithm that will use in poultry disease predictions. The clustering will assist in improving further analysis of the poultry symptoms data in detecting outliers. Analyzing outlier can reveal surprising facts hidden inside data like ambiguous patterns that are still assumed to belong to one of the predefined or undefined classes. Clustering is important in detecting outlier to avoid the high cost of misclassification. In order to cater for the complex nature of data of our problem domain, clustering technique based on machine learning approaches such as self organizing map (SOM), kernel machines, fuzzy methods, etc for clustering poultry symptoms (based on observation data – body, feathers, skin, head, muscle, lung, heart, intestines, ovary, etc) will prove to be a promising tool.

In this paper, a new two-level clustering algorithm is proposed. The idea is that the first level is to train the data by the SOM neural network and the clustering at the second level is a rough set based incremental clustering approach [9], which will be applied on the output of SOM and requires only a single neurons scan. The optimal number of clusters can be found by rough set theory which groups the given neurons into a set of overlapping clusters (clusters the mapped data respectively). Then the overlapped neurons will be assigned to the true clusters they belong to, by apply

genetic algorithm. A genetic algorithm has been adopted to minimize the uncertainty that comes from some clustering operations. In our previous work [3] the hybrid SOM and rough set has been applied to catch the involved ambiguity of clusters but the experiment results show that the proposed algorithm (Genetic Rough SOM) outperforms the previous one. the next important process is to collect poultry data from the common and important diseases, which can affect the respiratory and non-respiratory system of poultry. The first phase is to identify the format and values for input parameters from available information. The second phase is to investigate and develop data conversion and reduction algorithms for input parameters.

This paper is organized as following; in section II the basics of SOM algorithm are outlined. The basic of rough set incremental clustering approach are described in section III. In section IV the essence of genetic algorithm is described. The proposed algorithm is presented in section V. Section VI is dedicated to experiment results and section VII provides brief conclusion and future works.

II. SELF ORGANIZING MAP

Competitive learning is an adaptive process in which the neurons in a neural network gradually become sensitive to different input categories, sets of samples in a specific domain of the input space. A division of neural nodes emerges in the network to represent different patterns of the inputs after training.

The division is enforced by competition among the neurons: when an input x arrives, the neuron that is best able to represent it wins the competition and is allowed to learn it even better. If there exist an ordering between the neurons, i.e. the neurons are located on a discrete lattice, the competitive learning algorithm can be generalized. Not only the winning neuron but also its neighboring neurons on the lattice are allowed to learn, the whole effect is that the final map becomes an ordered map in the input space. This is the essence of the SOM algorithm. The SOM consist of m neurons located on a regular low-dimensional grid, usually one or two dimensional. The lattice of the grid is either hexagonal or rectangular.

The basic SOM algorithm is iterative. Each neuron i has a d -dimensional feature vector $w_i = [w_{i1}, ..., w_{id}]$. At each training step t, a sample data vector $x(t)$ is randomly chosen for the training set. Distance between $x(t)$ and all feature vectors are computed. The winning neuron, denoted by c, is the neuron with the feature vector closest to $x(t)$:

$$c = \arg\min_i \|x(t) - w_i\|, \qquad i \in \{1, ..., m\} \tag{1}$$

A set of neighboring nodes of the winning node is denoted as N_c. We define $h_{ic}(t)$ as the neighborhood kernel function around the winning neuron c at time t. The neighborhood kernel function is a non-increasing function of time and of the distance of neuron i from the winning neuron c. The kernel can be taken as a Gaussian function:

$$h_{ic}(t) = e^{-\frac{\|Pos_i - Pos_c\|^2}{2\sigma(t)^2}} \tag{2}$$

where Pos_i is the coordinates of neuron i on the output grid and $\sigma(t)$ is kernel width. The weight update rule in the sequential SOM algorithm can be written as:

$$w_i(t+1) = \begin{cases} w_i(t) + \varepsilon(t) h_{ic}(t)\left(x(t) - w_i(t)\right) \forall i \in N_c \\ w_i(t) \qquad\qquad ow \end{cases} \tag{3}$$

Both learning rate $\varepsilon(t)$ and neighborhood $\sigma(t)$ decrease monotonically with time. During training, the SOM behaves like a flexible net that fold onto a cloud formed by training data. Because of the neighborhood relations, neighboring neurons are pulled to the same direction, and thus feature vectors of neighboring neurons resemble each other. There are many variants of the SOM [10, 11]. However, these variants are not considered in this paper because the proposed algorithm is based on SOM, but not a new variant of SOM.

The 2D map can be easily visualized and thus give people useful information about the input data. The usual way to display the cluster structure of the data is to use a distance matrix, such as U-matrix [12]. U-matrix method displays the SOM grid according to neighboring neurons. Clusters can be identified in low inter-neuron distances and borders are identified in high inter-neuron distances. Another method of visualizing cluster structure is to assign the input data to their nearest neurons. Some neurons then have no input data assigned to them. These neurons can be used as the border of clusters [13].

III. ROUGH SET INCREMENTAL CLUSTERING

This algorithm is a soft clustering method employing rough set theory [14]. It groups the given data set into a set of overlapping clusters. Each cluster is represented by a *lower approximation* and an *upper approximation* $(\underline{A}(C), \overline{A}(C))$ for every cluster $C \subseteq U$. Here U is a set of all objects under exploration. However, the lower and upper approximations of $C_i \in U$ are required to follow some of the basic rough set properties such as:

(1) $\emptyset \subseteq \underline{A}(C_i) \subseteq \overline{A}(C_i) \subseteq U$

(2) $\underline{A}(C_i) \cap \underline{A}(C_j) = \emptyset,\ i \neq j$

(3) $\underline{A}(C_i) \cap \overline{A}(C_j) = \emptyset,\ i \neq j$

(4) If an object $u_k \in U$ is not part of any lower approximation, then it must belong to two or more upper approximations.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Note that (1)-(4) are not independent. However enumerating them will be helpful in understanding the basic of rough set theory.

The lower approximation $\underline{A}(C)$ contains all the patterns that definitely belong to the cluster C and the upper approximation $\overline{A}(C)$ permits overlap. Since the upper approximation permits overlaps, each set of data points that are shared by a group of clusters define *indiscernible set*. Thus, the ambiguity in assigning a pattern to a cluster is captured using the upper approximation. Employing rough set theory, the proposed clustering scheme generates soft clusters (clusters with permitted overlap in upper approximation).

For a rough set clustering scheme and given two objects $u_h, u_k \in U$ we have three distinct possibilities:

- Both u_k and u_h are in the same lower approximation $\underline{A}(C)$.
- Object u_k is in lower approximation $\underline{A}(C)$ and u_h is in the corresponding upper approximation $\overline{A}(C)$, and case 1 is not applicable.
- Both u_k and u_h are in the same upper approximation $\overline{A}(C)$, and case 1 and 2 are not applicable.

The quality of a conventional clustering scheme is determined using within-group-error [15] Δ given by:

$$\Delta = \sum_{i=1}^{m} \sum_{u_h, u_k \in C_i} distance\ (u_h, u_k) \tag{4}$$

where u_h, u_k *are objects in the same cluster* C_i.

For the above rough set possibilities, three types of equation (4) could be defined as following:

$$\begin{aligned}
\Delta_1 &= \sum_{i=1}^{m} \sum_{u_h, u_k \in \underline{A}(X_i)} distance\ (u_h, u_k) \\
\Delta_2 &= \sum_{i=1}^{m} \sum_{u_h \in \underline{A}(X_i)\ and\ u_k \in \overline{A}(X_i)} distance\ (u_h, u_k) \\
\Delta_3 &= \sum_{i=1}^{m} \sum_{u_h, u_k \in \overline{A}(X_i)} distance\ (u_h, u_k)
\end{aligned} \tag{5}$$

The total error of rough set clustering will then be a weighted sum of these errors:

$$\Delta_{total} = w_1 \times \Delta_1 + w_2 \times \Delta_2 + w_3 \times \Delta_3 \ where\ \ w_1 > w_2 > w_3. \tag{6}$$

Since Δ_1 corresponds to situations where both objects definitely belong to the same cluster, the weight w_1 should have the highest value.

IV. Genetic Algorithm

Genetic algorithm was proposed by John Holland in early 1970s, it applies some of natural evolution mechanism such as *crossover*, *mutation*, and survival of the fitness to optimization and machine learning. GA provides very efficient search method working on population, and has been applied to many problems of optimization and classification [16]-[17]. General GA process is as follows:

(1) Initial the population of genes.
(2) Calculates the fitness of each individual in the population.
(3) Reproduce the individual selected to form a new population according to each individual's fitness.
(4) Perform crossover and mutation on the population.
(5) Repeat step (2) through (4) until some condition is satisfied.

Crossover operation swaps some part of genetic bit string within parents. It emulates just as crossover of genes in real world that descendants are inherited characteristics from both parents. Mutation operation inverts some bits from whole bit string at very low rate. In real world we can see that some mutants come out rarely. Fig.1 shows the way of applying crossover and mutation operations to genetic algorithm. Each individual in the population evolves to getting higher fitness generation by generation.

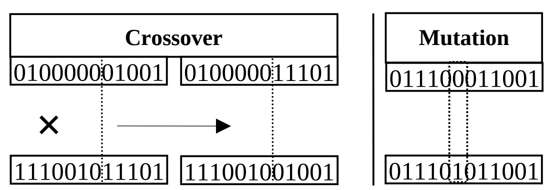


Figure 1. Crossover and Mutation

V. Genetic Rough set Clustering of the Self Organizing Map

In this paper rectangular grid is used for the SOM. Before training process begins, the input data will be normalized. This will prevent one attribute from overpowering in clustering criterion. The normalization of the new pattern $X_i = \{x_{i1}, ..., x_{id}\}$ for $i = 1,2,..., N$ is as following:

$$X_i = \frac{X_i}{\|X_i\|}. \tag{7}$$

Once the training phase of the SOM neural network completed, the output grid of neurons which is now stable to network iteration, will be clustered by applying the rough set algorithm as described in the previous section. The similarity measure used for rough set clustering of neurons is *Euclidean distance* (the same used for training the SOM). In this proposed method

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

(see Fig.2) some neurons, those never mapped any data are excluded from being processed by rough set algorithm.

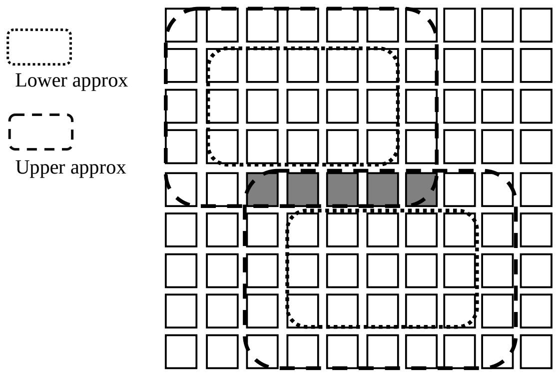


Figure 2. Clustering of the Self Organizing Map. The overlapped neurons are highlited for two clusters.

From the rough set algorithm it can be observed that if two neurons are defined as indiscernible (those neurons in the upper approximation of two or more clusters), there is a certain level of similarity they have with respect to the clusters they belong to and that similarity relation has to be symmetric. Thus, the similarity measure must be symmetric.

According to the rough set clustering of the SOM, overlapped neurons and respectively overlapped data (those data in the upper approximation) are detected. In the experiments, to calculate errors and uncertainty, the previous equations will be applied to the results of SOM (clustered and overlapped data). Then for each overlapped neuron a gene is generated that represents the alternative distances from each cluster leader. Fig.3 shows an example of the generated genes for m overlapped neurons on n existing cluster leaders.

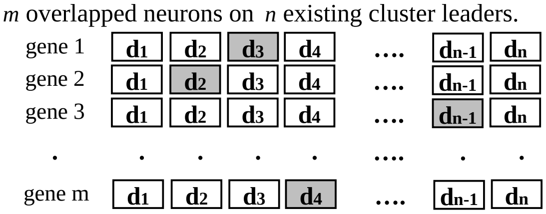


Figure 3. Generated genes. *m* number of overlapped neurons and *n* is number of existing clusters. The highlighted d_i is the optimize one that minize the fitness function

After the genes have been generated the genetic algorithm is employed to *minimize* the following *fitness function* which represents the total sum of each d_j of the related gene:

$$F = \sum_{i=1}^{m} \sum_{j=1}^{n} g_i(d_j) \tag{8}$$

The aim of the proposed approach is making the genetic rough set clustering of the SOM to be as precise as possible. Therefore, a precision measure needs to be used for evaluating the quality of the proposed approach. A possible precision measure can be defined as the following equation [14]:

$$certainty = \frac{\text{Number of objects in lower approx}}{\text{Total number of objects}} \tag{9}$$

VI. Experiment Results

To demonstrate the effectiveness of the proposed clustering algorithm GR-SOM (Genetic Rough set Incremental clustering of the SOM), two phases of experiments has been done on the well known Iris data set [18] and our gathered data. The Iris data set, which has been widely used in pattern classification, consists of 150 data points of four dimensions and our collected data has 48 data points. The Iris data are divided into three classes with 50 points each. The first class of Iris plant is linearly separable from the other two. The other two classes are overlapped to some extent.

The first phase of experiments, presents the uncertainty that comes from the data set and in the second phase the errors has been generated. The results of GR-SOM and RI-SOM [3] (Rough set Incremental SOM) are compared to I-SOM [4] (Incremental clustering of SOM). The input data are normalized such that the value of each datum in each dimension lies in $[0,1]$.

For training, SOM 10×10 with 100 epochs on the input data is used. The general parameters for the genetic algorithm have been configured as Table I. Fig.4 shows the certainty generated from epoch 100 to 500 by (9) on the mentioned data set. From the gained certainty it's obvious that the GR-SOM could efficiently detect the overlapped data that have been mapped by overlapped neurons (table II).

In the second phase, the same initialization for the SOM has been used. The errors that come from the data sets, according to the (5) and (6) have been generated by our proposed algorithms (table III). The weighted sum (6) has been configured as (10).

TABLE I. General parameters of the Genetic Algorithm

Population Size	50
Number of Evaluation	10
Crossover Rate	0.25
Mutation Rate	0.001
Number of Generation	100

TABLE II. The certainty-level of GR-SOM, RI-SOM and I-SOM on the Iris data set from epoch 100 to 500.

Epoch	***100***	***200***	***300***	***400***	***500***
I-SOM	33.33	65.23	76.01	89.47	92.01
RI-SOM	67.07	73.02	81.98	91.23	**97.33**
GR-SOM	69.45	74.34	83.67	94.49	**98.01**

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

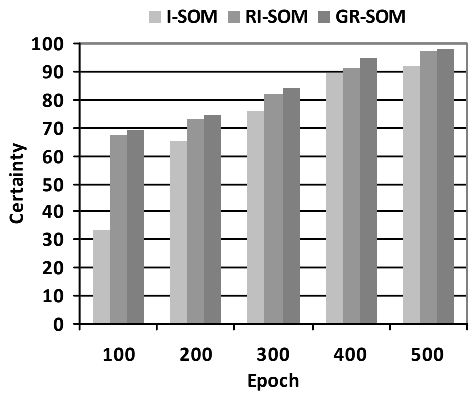


Figure 4. Comparison of the certainty-level of GR-SOM, RI-SOM and I-SOM on the Iris data set.

$$\sum_{i=1}^{3} w_i = 1$$
$$\textit{and for each } w_i \textit{ we have}: \quad (10)$$
$$w_i = \frac{1}{6} \times (4-i).$$

TABLE III. COMPARATIVE GENERATED ERRORS OF GR-SOM AND I-SOM ON THE IRIS DATA SET ACCORDING TO EQUATIONS (5) AND (6).

	Method	Δ_1	Δ_2	Δ_3	Δ_{total}
Iris Data set	**GR-SOM**	1.05	0.85	0.04	**1.4**
	I-SOM				2.8

Furthermore, to demonstrate the effectiveness of the proposed clustering algorithm (RI-SOM), two data sets, one artificial and one real word data set were used in our experiments. The results are compared to I-SOM (Incremental clustering of SOM). The input data are normalized such that the value of each datum in each dimension lies in $[0,1]$. For training SOM 10×10 with 100 epochs on the input data is used.

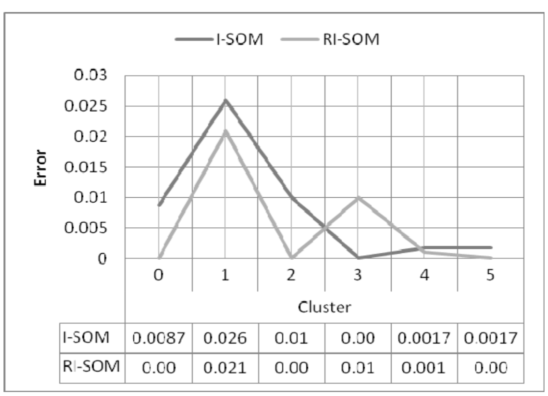


Figure 5. Comparison the error between I-SOM and RI-SOM proposed algorithms on artificial data set.

The artificial data set is a 569 30-dimensional data set which is trained twice, once with I-SOM and once with RI-SOM. The errors of generated results are calculated from the difference between the results of equation (9) and 1, see “Fig. 5”.

From the “Fig. 5” it could be observed that the proposed RI-SOM algorithm generates less error in cluster prediction compare to I-SOM.

VII. CONCLUSION AND FUTURE WORK

In this paper a two-level based clustering approach (GR-SOM), has been proposed to predict clusters of high dimensional data and to detect the uncertainty that comes from the overlapping data. The approach is based on the rough set theory that employs a soft clustering which can detects overlapped data from the data set and makes clustering as precise as possible, then GA is applied to find the true cluster for each overlapped data. The results of the both phases indicate that GR-SOM is more accurate and generates fewer errors as compare to crisp clustering (I-SOM).

The proposed algorithm detects accurate overlapping clusters in clustering operations. As the future work, the overlapped data also could be assigned correctly to true clusters they belong to, by assigning *fuzzy membership value* to the indiscernible set of data. Also a weight can be assigned to the data’s dimension to improve the overall accuracy.

REFERENCES

[1] T. Kohonen, “Self-organized formation of topologically correct feature maps”, *Biol.* Cybern. 43. 1982, pp. 59–69.

[2] T. Kohonen, Self-Organizing Maps, Springer, Berlin, Germany, 1997.

[3] M.N.M Sap and Ehsan Mohebi, “Hybrid Self Organizing Map for Overlapping Custers”. *The Springer-Verlag Proceedings of the CCIS 2008.* Hainan Island, China. Accepted.

[4] M.N.M Sap and Ehsan Mohebi, “Rough set Based Clustering of the Self Organizing Map”. *The IEEE Computer Scociety Proceeding of the 1st Aseian Conference on Intelligent Information and Database Systems 2009.* Dong Hoi, Vietnam. Accepted

[5] M.N.M Sap and Ehsan Mohebi, “A Novel Clustering of the SOM using Rough set”. *The IEEE Proceeding of the 6th Student Conference on Research and Development 2008”*. Johor, Malaysia 2008. Accepted

[6] R.M. Gray., “Vector quantization”. *IEEE Acoust. Speech, Signal Process. Mag*. 1 (2) 1984. pp. 4–29.

[7] N.R. Pal, J.C. Bezdek, and E.C.K. Tsao, “Generalized clustering networks and Kohonen’s self-organizing scheme”. *IEEE Trans. Neural Networks* (4) 1993. pp. 549–557.

[8] J. Han, M. Kamber, “Data mining: concepts and techniques”, Morgan-Kaufman, San Francisco, 2000.

[9] S. Asharaf, M. Narasimha Murty, and S.K. Shevade, “Rough set based incremental clustering of interval data”, *Pattern Recognition Letters*, Vol. 27, 2006, pp. 515-519.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[10] Yan and Yaoguang., "Research and application of SOM neural network which based on kernel function". *Proceeding of ICNN&B'05*. Vol.1, 2005. pp. 509- 511.

[11] M.N.M. Sap and Ehsan Mohebi. "Outlier Detection Methodologies: A Review". Journal of Information Technology, UTM, Vol. 20, Issue 1, 2008. pp. 87-105.

[12] A. Ultsch, H.P. Siemon., "Kohonen's self organizing feature maps for exploratory data analysis". *Proceedings of the International Neural Network Conference*, Dordrecht, Netherlands 1990. pp. 305–308.

[13] X. Zhang, Y. Li. "Self-organizing map as a new method for clustering and data analysis". *Proceedings of the International Joint Conference on Neural Networks*, Nagoya, Japan 1993. pp. 2448–2451.

[14] Pawlak, Z., "Rough sets". Internat. J. Computer Inf. Sci. vol.11, 1982. pp. 341–356.

[15] S.C. Sharma and A. Werner., "Improved method of grouping provincewide permanent traffic counters". *Transaction Research Report 815*, Washington D.C. 1981 pp. 13-18 .

[16] Goldberg D.E, "Genetic Algorithm in Search Optimization and Machine Learning". Addison-Wesley Pubishing Co.inc, 1989.

[17] Ebrehart, R, Simpson P. Dobbins R., "Comptational Intelligent PC Tools", Waite Group Press, 1996.

[18] UCIMachineLearning,www.ics.uci.edu/mlearn/MLRepository.html.

Secured Communication through Hybrid Crypto-Steganography

A. Joseph Raphael
Research Scholar – Karpagm University, Coimbatore, India and Lecturer in Information Technology, Ibra College of Technology, Sultanate of Oman
raphaelaj@gmail.com

Dr. V.Sundaram
Head and Director, Department of Computer Applications
Karpagam College of Engineering
Coimbatore, India
dr.vsundaram@gmail.com

***Abstract*-In this paper we present a hybrid technology of mixing cryptography and steganography to send secret messages. This method has got the advantages of both the methods, and even if one fails the other comes to the rescue. In cryptography we have used the RSA method for encryption and decryption of the original message and further the LSB (Least Significant Bit) method is used to hide the encrypted message in the cover image and send to the recipient. The original message is retrieved by the reverse process, first by collecting the LSB of the pixels and then by RSA decryption. Since the private key for RSA method is very difficult to find, this method we suggested is a strong encryption method and the messages can be communicated in much secured way in an insecure channel.**

Keywords-stegano object; cryptosystem

I. INTRODUCTION

The importance of secret communication by governments and private organizations has increased a lot. By the advent of e-commerce and the increasing trade, administration, terrorism, security of nations, secret communications have taken the top most priority in the communication sector.

Steganography is derived from the Greek words stegos, meaning roof or covered and graphia which means writing, is the art and science of hiding or embedding a secret message over a piece of information such as an image, audio or video and sent them over an insecure channel to the recipient so that no one can detect or decode the secret message.

Steganography is very closely related to Cryptography, both are used to maintain the data in a confidential manner. The main difference between the two is that with Cryptography the message is scrambled and anybody can see that both parties are communicating in secret. Steganography on the other hand, hides the existence of a secret message and in the best case nobody can see that both parties are communicating in secret. In some situations, sending an encrypted message will arouse suspicion while an "invisible" message will not do so. Both sciences can be combined to produce better protection of the message. In this case, when the steganography fails and the message can be detected, it is still of no use as it is encrypted using cryptography techniques. Moreover, to avoid unnecessary suspicious the messages can be hidden in the cover image which can be selected from seasonal greetings.

The power of a cryptographic / steganographic system should depend only on a small part of information namely the key to uncover the cipher text/material.

We employ RSA public key cryptography for the encryption and decryption of the original message followed by steganography using LSB. In both methods the encryption and decryption depend only on a small key.

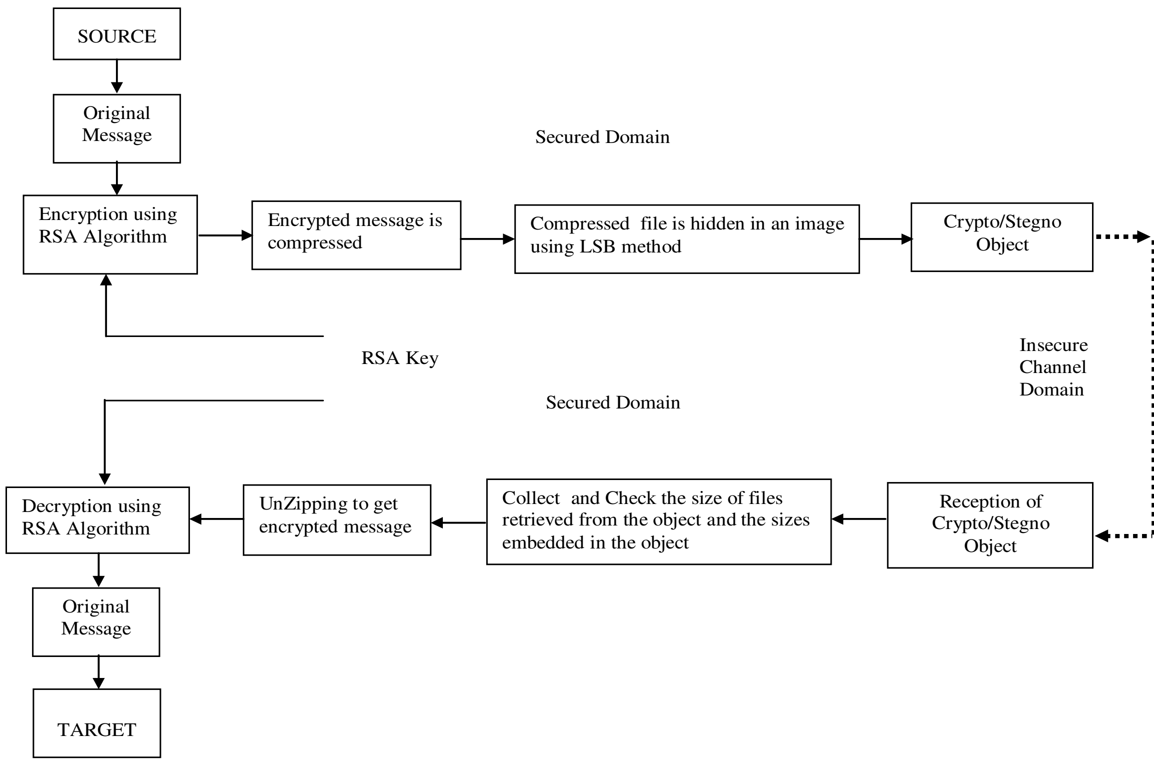
A digital key is a set of bits that are employed to encrypt and decrypt the messages. A public key cryptography uses two different keys. 1) A public key to encrypt the original message. 2) A private key to decrypt the cipher text and expose the original message.

Public and Private keys are generated in pairs so that only a specific pair of keys can perform encryption and decryption. The public key is made known to everyone whereas the matching private key is kept as a secret by the owner. During mid 1970s cryptogists introduces the concept of public key cryptosystems. In this system everyone can have a public key for encryption; however the decryption keys are kept secret only by the intended recipient who can decrypt it.

In 1976 Ronald Rivest, AdlShamir, and Leonard Adleman introduced a public key cryptosystem known as RSA system [1]. This system is based on the modular exponentiation modulo, the product of the large primes. The encryption key consists of a modulus $n = pq$ where p and q are large primes more than 200 digits each and an exponent e that is relatively prime to $(p-1)(q-1)$. The two large primes p and q can be found quietly on a computer using probabilistic primality tests. The product of the primes $n = pq$ with about 400 digits cannot be factored in a reasonable length of time and hence the decryption of the cipher text cannot be done quickly without a separate decryption key.

Here, we describe the process how the original message is encrypted, masked and sent, and on the other end how it is received, unmasked and decrypted into original message, and the same is presented in the schematic diagram below:

II. AFFINE TRANSFORMATION



At first the original message (OM) consisting of a sequence of letters is converted into another sequence of letters using a ceaser cipher or by a more general affine transformation defined by the bijection f.

$$f : c = f(p) = (ap + b) \bmod 26$$

Here p is the rank of a particular letter in OM which is converted to the number c. In the encrypted version of the message, the letter represented by p in OM is replaced with the letter represented by c.

III. RSA ENCRYPTION

The new message obtained by the affine transformation is encrypted by the RSA method as follows:

1. Each letter in the message is represented as its ASCII code number and all such ASCII code is converted into an equivalent binary number using 8 bits.
2. Then each pair of characters are grouped into blocks by taking two adjacent 8 bits side by side as one 16 bit number. If odd numbers of letters are present the binary number corresponding to the last character is padded with zeros in the beginning to have 16 bits.
3. Each of the message block is represented as an equivalent decimal number set that will be encrypted using the RSA rule or formula $(block)^E \bmod n$ to get a set of encrypted integers [1]. This encryption is done by using the following modular exponentiation algorithm.

Procedure for modular exponentiation

b:integer, $n = (a_{k-1}a_{k-2}....a_1a_0)$,
m: positive integers
x:=1
power:=b mod m
for i = 0 to k-1
begin
  if a_i=1 then x:=(x.power) mod m
  power := (power.power) mod m
end
{ x equals b^n mod m }

4. These encrypted blocks are converted into a sequence of 16 bit binary numbers that is split into two 8 bit numbers.

Usually the encrypted message is directly inserted into an image using LSB method, which generally requires a lot of space to hide a relatively few bits of information which is also one of the disadvantages of LSB method. To overcome this drawback, the encrypted message is first compressed before it is embedded so that a large amount of information can then be hidden. Using the Least significant bit insertion method of Steganography the

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

binary equivalent of the compressed file is encoded into the Least significant bit of each byte as each pixel is represented by 3 bytes of a 24 bit image.

IV. USE OF DIGITAL IMAGES

Most popular file formats being BMP (Bitmaps) GIF (Graphics Interchange Format) and JPEG (Joint Photographic experts Group). Of these formats the first two provide a loseless message transfer where the last method provides a lossy transfer. Loseless files won't shatter the image much if the pixel intensities are altered, whereas the Lossy files tend to shatter heavily while the pixel values are altered. Hence the lossy images are preferred much for data transfer.

Images are classified as 8 bit or 24 bit images based upon the number of bits they use to represent a color. The image files supporting larger intensities are preferred for covert transmission and are stored in variety of file formats. Image file is a large array of pixel intensities (Color Values). These pixel values form the raster data of the image and could be used to impose the message. Each one of these pixels has its own color, and it is represented internally as separate quantities of red, green and blue. Each of these color levels may range between 0 (none of the color) and 255 (a full amount of the color). A pixel with an RGB value of (0,0,0) is black, and one with a value of (255,255,255) is white. For a 24 bit image this is simple because 24 bit images are stored internally as RGB triples, and all that needs to be done is to spread the bits and save out the new file. The images are also very large as they contain 3 bytes for every pixel (for a 640 x 480 image this is 640 x 480 x 3 = 921600 bytes).

When applying LSB techniques to each byte of a 24-bit image, three bits can be encoded into each pixel. (As each pixel is represented by three bytes.) Any changes in the pixel bits will be indiscernible to the human eye [2]-[3]. For example, the letter A can be hidden in three pixels. Assume the original three pixels are represented by the three 24-bit words as below

(00100111 11101001 11001000)

(00100111 11001000 11101001)

(11001000 00100111 11101001)

The binary value for the letter A is **10000011**. Inserting the binary value for A into the three pixels, starting from the top left byte, would result in

(00100111 1110100**0** 11001000)

(0010011**0** 11001000 1110100**0**)

(11001000 00100111 11101001)

The emphasized bits are the only bits that are actually changed. The main advantage of LSB insertion is that data can be hidden in the least and second to least bits and still to the human eye, the resulting image with embedded data will look practically identical to the cover image. Notice that only the bolded bits had to be changed in order to create the letter A. On the average only less than 50% of the bits would have to be changed in an LSB encoding scheme.

The outcome of the above stage produces a Crypto/Stegno Object which consist of hidden message in compressed form, within a cover image. Also, the size of the encrypted message, the size of the compressed file and the size of the crypto/stegno object itself can be embedded in the same cover image after embedding the bits of the original compressed file. All the above said stages are in the Secured Domain, from where the Crypto/Stegno Object is passed to an Insecured Channel Domain to the recipient target. Once the crypto/stegno object is arrived at the recipient target, the following embedded informations are retrieved : 1) all the LSBs of the pixels are collected to form the compressed file 2) size of the encrypted message and 3) size of the crypto/stegno object. To ensure that, the full and correct message is reached at the recipient target, the following comparisons are made between 1) the size of the crypto/stegno object received at the recipient target and the size of the crypto/stegno object which is embedded in the crypto/stegno object and sent 2) the size of the compressed file obtained from the object and the size of the compressed file which is embedded in the object and dispatched 3) the size of the encrypted message after unzipping and the size of the encrypted message which is embedded in the object. If any mismatch occurs in any one of the sizes then it is understood that the crypto/stegno object is subjected to natural attack while passing over an insecured channel domain and the process has to start from the beginning.

On the other hand, if there is no mismatch, then from the received crypto/stegno object all the Least Significant Bits of the pixels are combined together to form a compressed file. Futher, the compressed file is unzipped to get an encrypted message from which the original message is obtained by undergoing the process of decryption using RSA Algorithm.

www.stegoarchive.com [4] is a website which have many stegno tools to automate the process of changing the LSB to allow for the insertion of some other data on an image and getting back the data and the image separately at the other end. A few of the more popular applications are ExStego, Jstego and hide4pgp. S-Tools is another tool that uses a different method for utilizing the LSB theory. It closely approximates the cover image and that could possibly cause extreme palette changes in the original image.

V. THE RSA DECRYPTION

The RSA decryption key D is the inverse E modulo of $(p-1)(q-1)$.

ie. $ED = 1 \bmod (p-1)(q-1)$, this can be found by using the euclidean algorithm. Then the decryption is done by the rule $(block)^D \bmod n$ where *block* is the cipher text.

The whole strength of the RSA method lies in the fact that even though $n = pq$ is known, the factorization of n is the most difficult problem (as against finding large primes

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

p and q). It is known that even most efficient factorization methods (till recently by 2002) requires billions of years to factor 400 digit integers. Hence when p and q are 200 digits primes, messages encrypted using $n = pq$ as the modulus cannot be found in a reasonable time unless p or q is known.

VI. CONCLUSION

The RSA-LSB crypto/stegno method suggested has sound mathematical and logical support. The decryption key for the methods depend only on a small part of information namely

i. for RSA – the exponent E and its inverse D

ii. for LSB, the collection of least significant digits.

The method is highly secure and decryption cannot be obtained by illegal persons within a reasonable period of time. For both the methods there are several application/software available and hence execution of the method is readily possible.

REFERENCES

[1] Kenneth H. Rosen, "Discrete Mathematics and its Applications", McGraw Hill, Fifth Edition.

[2] Neil F.Johnson and Sushil Jajodia, "Exploring Steganography: Seeing the Unseen", George Mason University, Available online at http://www.jjtc.com/pub/r2026.pdf.

[3] J.R. Krenn, "Steganography and steganalysis", Available at http://www.krenn.nl/univ/cry/steg/article.pdf

[4] James C.Judge, "Steganography Past, Present, Future", Avaiable at http://www.sans.org/reading_room/whitepapers /stenganography/steganography-past-present-future_552 .

AUTHORS PROFILE

A. Joseph Raphael obtained his Master degree in Computer Science from St. Joseph's College, Tiruchirapalli and Master of Philosophy in Computer Science from Alagappa University, Karaikudi. Currently, he is a PhD research scholar at Karpagam University, Coimbatore, India and also working as a lecturer in the department of Information Technology, Ibra College of Technology, Sultanate of Oman.

Dr. V. Sundaram earned his PhD in mathematics from Madras University. He is a research guide of Anna University, Coimbatore and Karpagam University in the field of computer science and computer applications. He is currently guiding several PhD students in the areas of theoretical computer science, network security, cryptography and data mining. He has published several papers in national and international journals and organized 5 national conferences. He is a life member of ISTE and member of Computer Society of India

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Lossy audio coding flowchart based on adaptive time- frequency mapping, wavelet coefficients quantization and SNR psychoacoustic output

Khalil Abid
Laboratory of Systems and Signal Processing (LSTS)
National Engineering School of Tunis (ENIT)
BP 37, Le Belvédère 1002, Tunis, Tunisia
Khalilabid06@yahoo.fr

Kais Ouni and Noureddine Ellouze
Laboratory of Systems and Signal Processing (LSTS)
National Engineering School of Tunis (ENIT)
BP 37, Le Belvédère 1002, Tunis, Tunisia

Abstract—**This paper describes a novel wavelet based audio synthesis and coding method. The adaptive wavelet transform selection and the coefficient bit allocation procedures are designed to take advantage of the masking effect in human hearing. They minimize the number of bits required to represent each frame of audio material at a fixed distortion level. This model incorporates psychoacoustic model into adaptive wavelet packet scheme to achieve perceptually transparent compression of high-quality audio signals.**

Keywords- D.W.T; Psychoacoustic Model; Signal to Noise Ratio; Quantization

I. INTRODUCTION

The vast majority of audio data on the Internet is compressed using some form of lossy coding, including the extremely popular MPEG1 Layer III (MP3) [1], Windows Media Archive (WMA) and Real Media (RM) formats. These algorithms can generally achieve compression ratios by using a combination of signal processing techniques, psychoacoustics and entropy coding,. most popular attention has been focused on lossy compression schemes like MP3, WMA and Ogg Vorbis. In general, these schemes perform some variant of either the Fast Fourier Transform (FFT) or Discrete Cosine Transformation (DCT) [8] to get a frequency-based representation of the sound waveform. Lossy algorithms generally take advantage of a branch of psychophysiology known as psychoacoustics that describes the ways in which humans perceive sound. By removing tones and frequencies that humans should not be able to hear, lossy algorithms can greatly simplify the nature of the data which they need to encode. By removing excess minor frequencies, the frequency representation of the sound data can now be efficiently compressed using any number of entropy coding techniques.

The wavelet transform becomes an emerging signal processing technique [13] and it is used to decompose and reconstruct non-stationary signals efficiently. The audio signal is non-periodic and it varies temporally. The wavelet transform can be used to represent audio signals [14] by using the translated and scaled mother wavelets, which are capable to provide multi-resolution of the audio signal. This property of wavelet can be used to compress audio signal. The DWT consists of banks of low pass filters, high pass filters and down sampling units. Half of the filter convolution results are discarded because of the down sampling at each DWT decomposition stage [6] [11]. Only the approximation part of the DWT wavelet results is kept so that the number of samples is reduced by half. The level of decomposition is limited by the distortion tolerable from the resulting audio signal.

II. STRUCTURE OF THE PROPOSED AUDIO CODEC

The main goal of this structure is to compress high quality audio maintaining transparent quality at low bit rates. In order to do this, the authors explored the usage of wavelets instead of the traditional Modified Discrete Cosine Transform (MDCT) [1]. Several steps are considered to achieve this goal:

- Design a wavelet representation for audio signals.
- Design a psychoacoustic model to perform perceptual coding and adapt it to the wavelet representation.
- Reduce the number of the non-zero coefficients of the wavelet representation and perform quantization over those coefficients.
- Perform extra compression to reduce redundancy over that representation
- Transmit or store the steam of data. Decode and reconstruct.
- Evaluate the quality of the compressed signal.
- Consider implementation issues.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

The flowchart of the proposed model is based on the following steps :

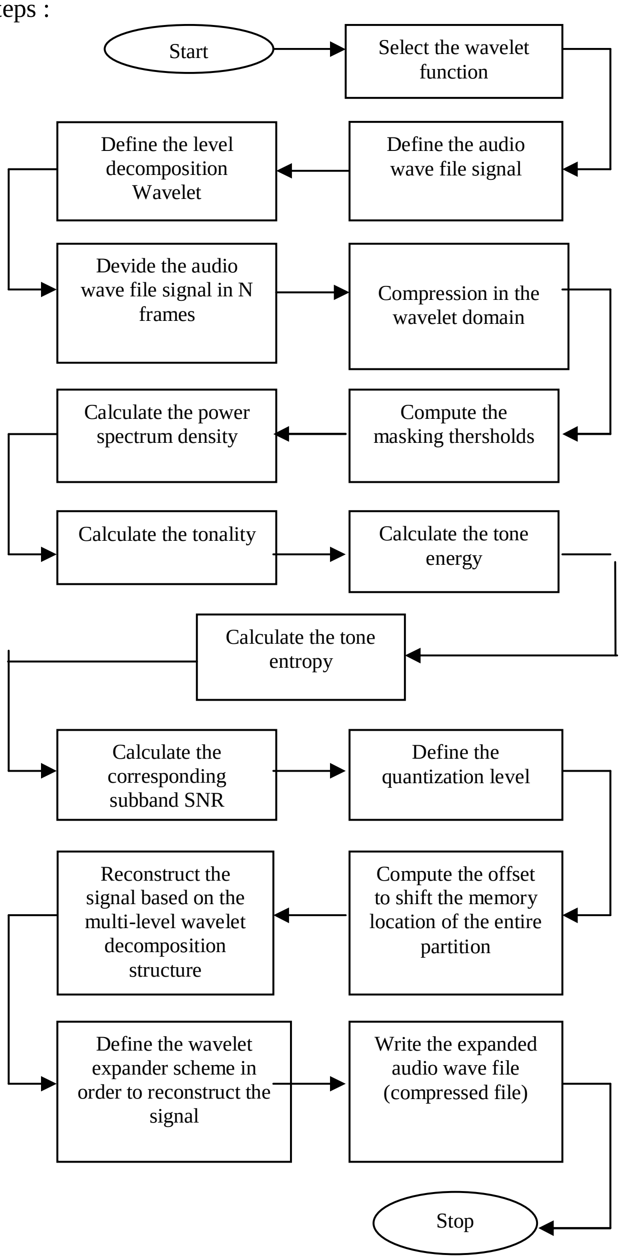


Figure 1. The different steps of the proposed audio wavelet compressed codec

The audio wave file is separated into small sections called frames (2048 samples). Each section is compressed using the proposed wavelet encoder and decoder. The encoder is consisting in four functional unit: the time to frequency mapping , the psychoacoustic model, the quantizer & coder and the frame packing unit. The function of the time to frequency mapping is used to decompose the input audio signal into multiple subbands for coding. This mapping is performed in three levels, labeled I ,II & III, which are caracterised with increasing complexity, delay and subjective performance. The algorithm in level I uses a band pass filter bank that devides the audio signal into 32 equal width subbands [4]. This filter bank is also found in level II and III. The design of this filter bank is a compromise between computational efficiency and perceptual performance. The algorithm in level II is a simple enhancement of level I; it improves compression performance by coding the audio data in larger groups. Finally the level III algorithm is much more refined in order to come closer the critical bands [2] [5] . The psychoacoustic model is key component in the encoder. Its function is to analyze the spectral content of the input audio signal by computing the signal to noise ratio for each subband. This information is used by the quantizer-coder to decide the available number of bits to quantize each subband. This dynamic allocation of bits is performed so as to minimize the audibility of quantization noise. Finally frame-packing unit assembles the quantized audio samples into decodable bit stream. The decoder consists of three functional units: the frame unpacking unit, the frequency sample reconstruction and the frequency to time mapping. The decoder simply reverses the signal processing operations performed in the encoder, converting the received stream of encoded bits into time domain audio signal.

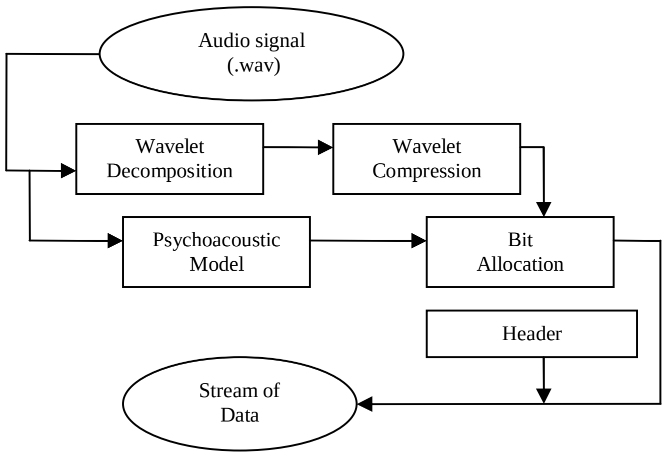


Figure 2. The audio waveet encoder

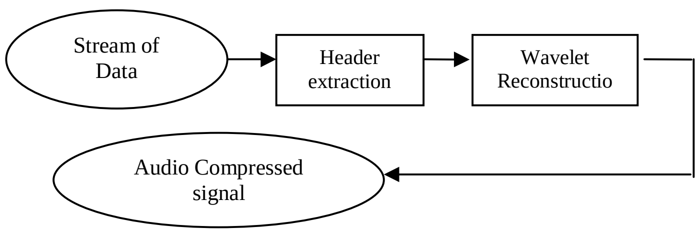


Figure 3. The audio wavelet decoder

III. THE PSYCHOACOUSTIC MODEL

The psychoacoustic model is a critical part of perceptual audio coding that exploits masking properties of the human auditory system. The psychoacoustic model analyzes signal content and combines induced masking curves to determine

what information below the masking threshold that is perceptually inaudible and should be removed. The psychoacoustic model is based on many studies of human perception. These studies have shown that the average human doesn't hear all frequencies the same. Effects due to different sounds in the environment and limitations of the human sensory system lead to facts that can be used to cut out unnecessary data in an audio signal. The two main properties of the human auditory system that make up the psychoacoustic model are the absolute threshold of hearing [1] [15] and the auditory masking [1]. Each one provides a way of determining which portions of a signal are inaudible and indiscernible to the average human, and can thus be removed from a signal.

A. The Absolute Threshold of Hearing

. To determine the effect of frequency on hearing ability, scientists played a sinusoidal tone at a very low power. The power was slowly raised until the subject could hear the tone. This level was the threshold at which the tone could be heard. The process was repeated for many frequencies in the human auditory range and with many subjects. As a result, the following plot was obtained. This experimental data can be modeled by the following equation, where f is frequency in Hertz [2]:

$$T_q(f) = 3.64\left(\frac{f}{1000}\right)^{-0.8} - 6.5e^{-0.6\left(\frac{f}{1000}-3.3\right)^2} + 0.001\left(\frac{f}{1000}\right)^4 \quad (1)$$

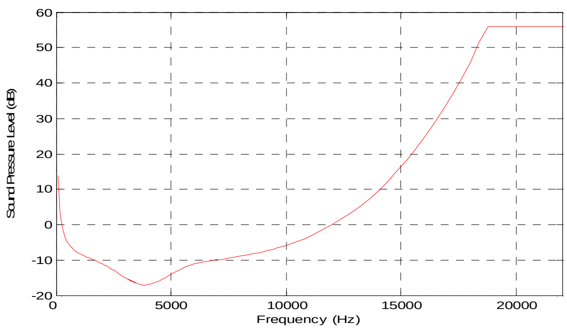


Figure 4. The absolute thershold of hearing

B. The Bark Frequency Scale

After many studies, scientists found that the frequency range from 20 Hz to 20000 Hz [3] [10] can be broken up into critical bandwidths [12], which are non-uniform, non-linear, and dependent on the heard sound. Signals within one critical bandwidth are hard to separate for a human observer [7]. A more uniform measure of frequency based on critical bandwidths is the Bark. From the earlier discussed observations, one would expect a Bark bandwidth to be smaller at low frequencies (in Hz) and larger at high ones. Indeed, this is the case. The Bark frequency scale can be approximated by the following equation [2]:

$$v(f) = 13artg(0.00076) + 3.5artg\left[\left(\frac{f}{7500}\right)^2\right] \quad (2)$$

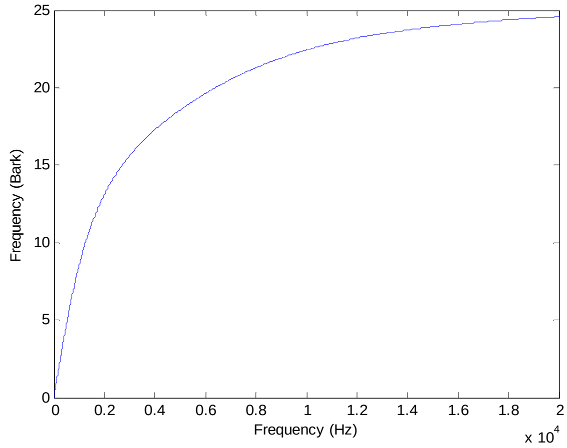


Figure 5. Relationship between Hertz and Bark Frequencies

C. Tone and Noise Masker Identification

Masking curces of tonal and noise maskers [1] have different shapes [1] therefore it is necessary to separate them.. To find tonal components it is necessary to find local maximas and then compare them with their neighbourhood components. This action hints Eq. 3 [1] [3]:

$$S_{SPL}(i) - S_{SPL}(i \pm \Delta_i) \geq 7 \quad (3)$$

where:

$$\Delta_i = +2 \quad \text{for} \quad i \in \left]2, 63\right[\quad (4)$$

$$\Delta_i = +2\,, +3 \quad \text{for} \quad i \in \left[63, 127\right[\quad (5)$$

$$\Delta_i = +2 \ldots +6 \quad \text{for} \quad i \in \left[127, 255\right[\quad (6)$$

$$\Delta_i = +2 \ldots +12 \quad \text{for} \quad i \in \left[255, 512\right[\quad (7)$$

According to ISO/IEC MPEG1, Psychoacoustic Analysis Model1 of MPEG1 audio standard [1] sound pressure level of the tonal masker is computed by Eq.8 as a summation of the spectral density of the masker and its neighbours:

$$X_{TM}(i) = 10.\log_{10}\left(\sum_{j=-1}^{1} 10^{\frac{S_{SPL}(i+j)}{10}}\right) \ [\text{dB}] \quad (8)$$

Sound Pressure level of the noise maskers is computed according to Eq. 9 as a summation of the sound pressure level of all spectral components in corresponding critical band.

$$X_{NM}(i) = 10.\log_{10}\left(\sum_{j=-1}^{1} 10^{\frac{S_{SPL}(i)}{10}}\right) \ [\text{dB}]\,, \quad y(i) \in b \quad (9)$$

where b represents the critical band, i index spectral components that lies in the corresponding critical band. Noise maskers are placed in the middle of the corresponding critical band.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

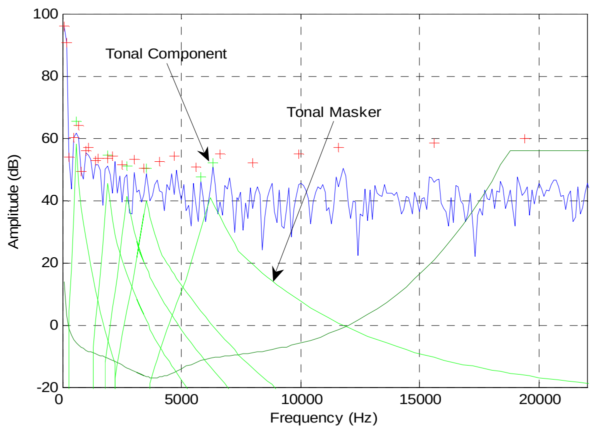


Figure 6. The Tonal Components and Tonal Maskers

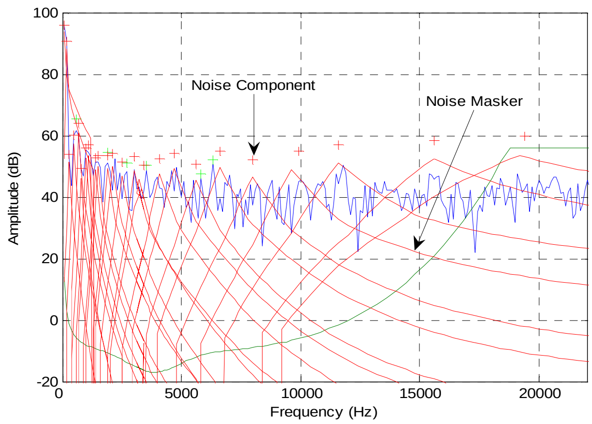


Figure 7. The Noise Components and Noise Masker

D. Masking Thershold Calculation

When tonal and noise maskers are identified, the masking thershold for each masker is determined. As defined in ISO/IEC MPEG1 Psychoacoustic Analysis Model 1 of MPEG1 audio standard [ ] tonal masker masking curve can be calculated the following equation 10 [1]:

$$M_{TM}(i,j) = X_{TM}(i) + MF(i,j) - 0.275.y(j) + 6.025 \qquad (10)$$

Where X_{TM} is a Sound Pressure Level of the tone masker. $y(j)$ is the masking curve position on the bark axis. $MF(i,j)$ is a masking function defined by Eq. 11. The constant 6.025 represents the top of the masking curve

$$MF(i,j) = 17.\Delta_y - 0.4X_{TM}(i) + 11 \qquad \Delta_y \in [-3,-1[\qquad (11)$$

$$MF(i,j) = (0.4X_{TM}(i) + 6).\Delta_y \qquad \Delta_y \in [-1,0[\qquad (12)$$

$$MF(i,j) = -17.\Delta_y \qquad \Delta_y \in [0,1[\qquad (13)$$

$$MF(i,j) = (1-\Delta_y).(17 - 0.15.X_{TM}(i)) - 17 \qquad \Delta_y \in [1,8[\qquad (14)$$

Identify applicable sponsor/s here. *(sponsors)*

Where $\Delta_y = y(i) - y(j)$ represents bark distance form the masker in barks.

Note : Outside the interval [-3,8[, MF is equal to $-\infty$

Masking curves of the noise maskers is defined by ISO/IEC MPEG1 Psychoacoustic Analysis Model 1 [1] and it is similar to the tone masker. The noise is defined by the following equation [1]:

$$M_{NM}(i,j) = X_{NM}(i) + MF(i,j) - 0.175.y(j) + 2.025 \qquad (15)$$

Where X_{NM} is a Sound Pressure Level of the noise masker. The constant 2.025 represents the top of the masking curve.

IV. BIT ALLOCATION

In order to determine the number of bits corresponding to each troncatured audio wave signal (2048 samples) we proceeded the following algorithm:

We start by listing all the tonal components characterized by the following condition [1] [9] :

$$X(i) > X(i-1) \;\&\; X(i) > X(i+1) \qquad (16)$$

Where $X(i)$ is the sound pressure level of the indexing (i) tonal component

For each tonal masker corresponding to the indexed (i) tonal component , we calculate the corresponding tone energy caracterized by the following equation:

$$E_{tm}(i) = 10.\log_{10}\left(\left(10^{\frac{X(i-1)}{10}}\right)^2 + \left(10^{\frac{X(i)}{10}}\right)^2 + \left(10^{\frac{X(i+1)}{10}}\right)^2\right) \qquad (17)$$

Then, we calculate the global energy of the all tones energie corresponding to the troncatured audio wave signal (2048 samples).

$$E_G = \sum_{i=1}^{N_{tm}} 10^{\frac{E_{tm}(i)}{10}} \qquad (18)$$

Note: N_{tm} is the total number of tonal maskers

All this allows to deduce the entropy using the following equation :

$$E = 10.\log_{10}\left(\frac{\sum_{i=1}^{N_{tm}} 10^{\frac{E_{tm}(i)}{10}}}{N_{tm}}\right) \qquad (19)$$

SNR is calculated using Eq.20 as a subtraction of the maximum of sound pressure level and the entropy:

$$SNR = Max(X) - E \quad \text{[dB]} \qquad (20)$$

Finally the number of bits corresponging to the troncatured signal is given by the following equation:

$$nb = \left(\frac{SNR}{6.02}\right) \qquad (21)$$

V. DIAGRAM OF THE WAVELET ENCODER AND DECODER

The flowchart of the wavelet codec is devided in 5 parts :

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

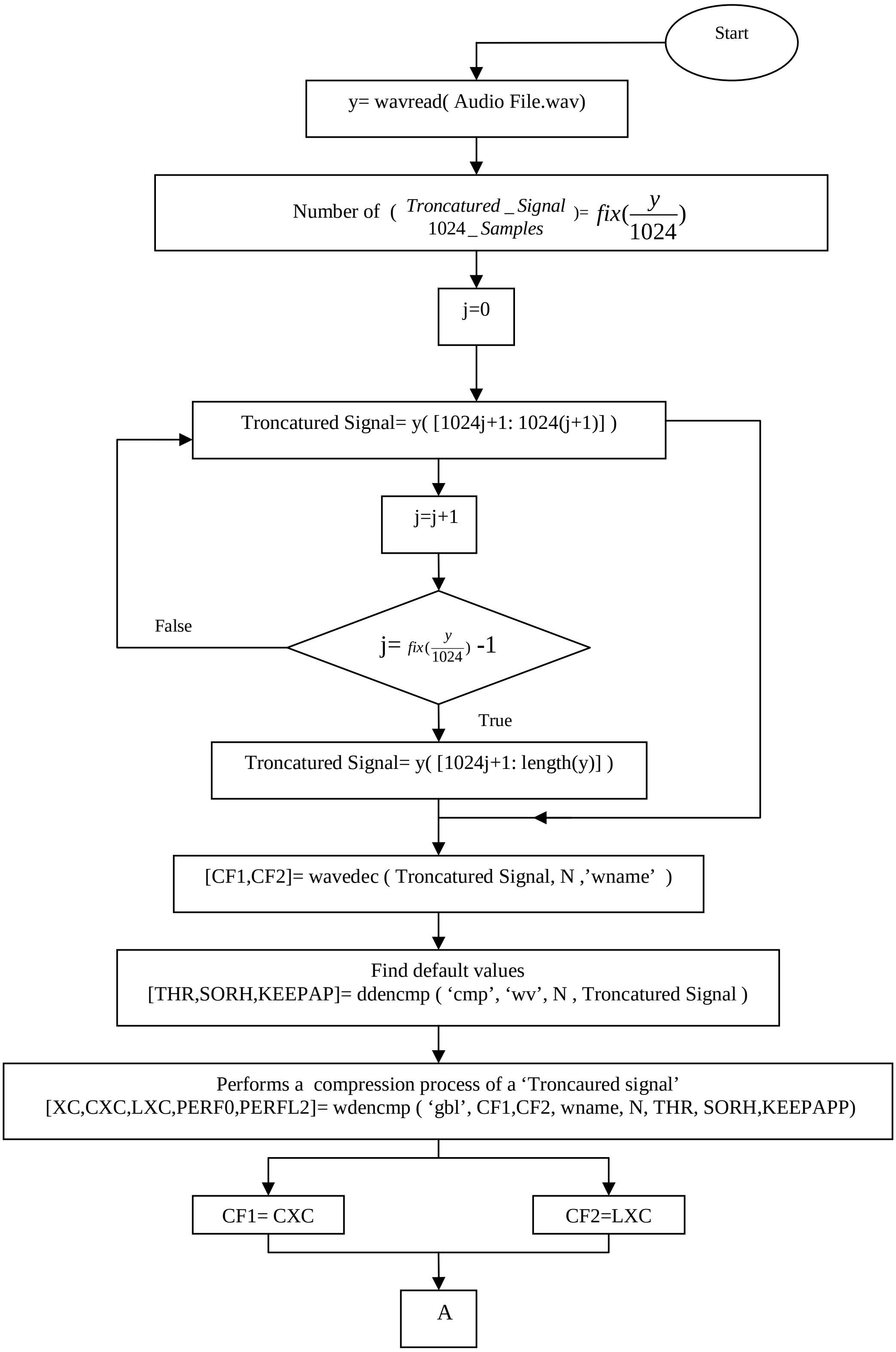


Figure 8. Diagram of the wavelet encoder and decoder (part 1)

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

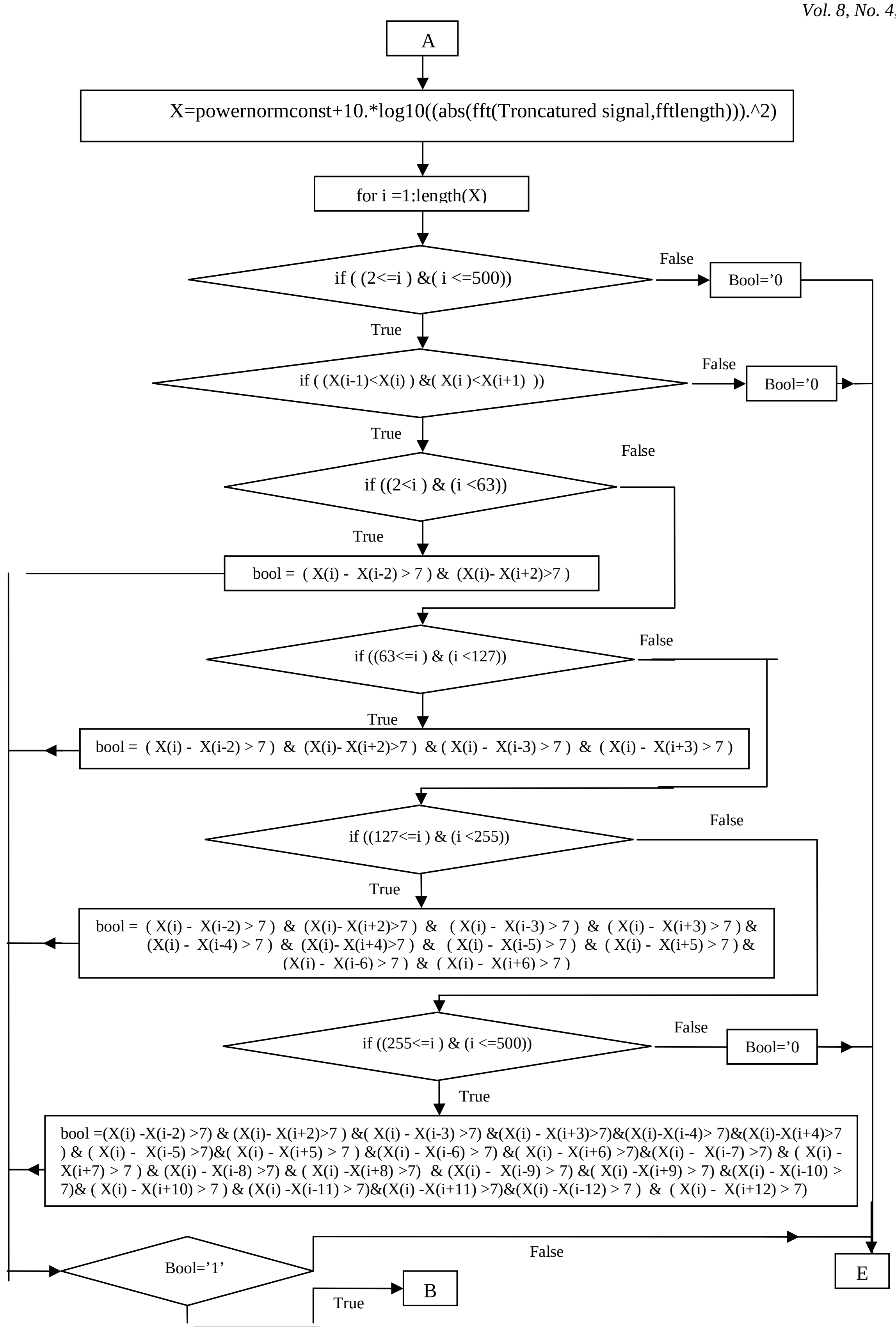


Figure 9. Diagram of the wavelet encoder and decoder (part 2)

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

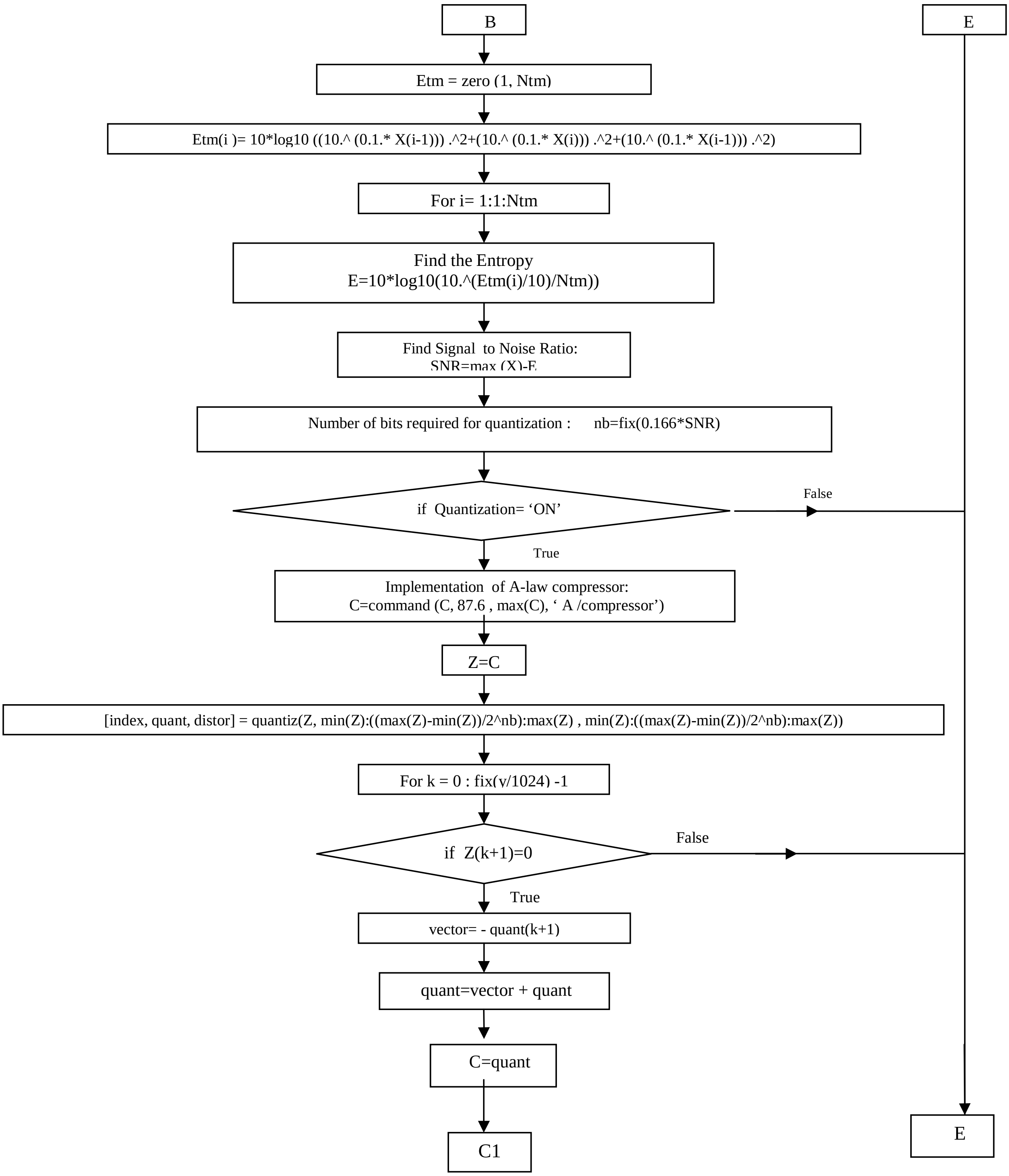


Figure 10. Diagram of the wavelet encoder and decoder (part 3)

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

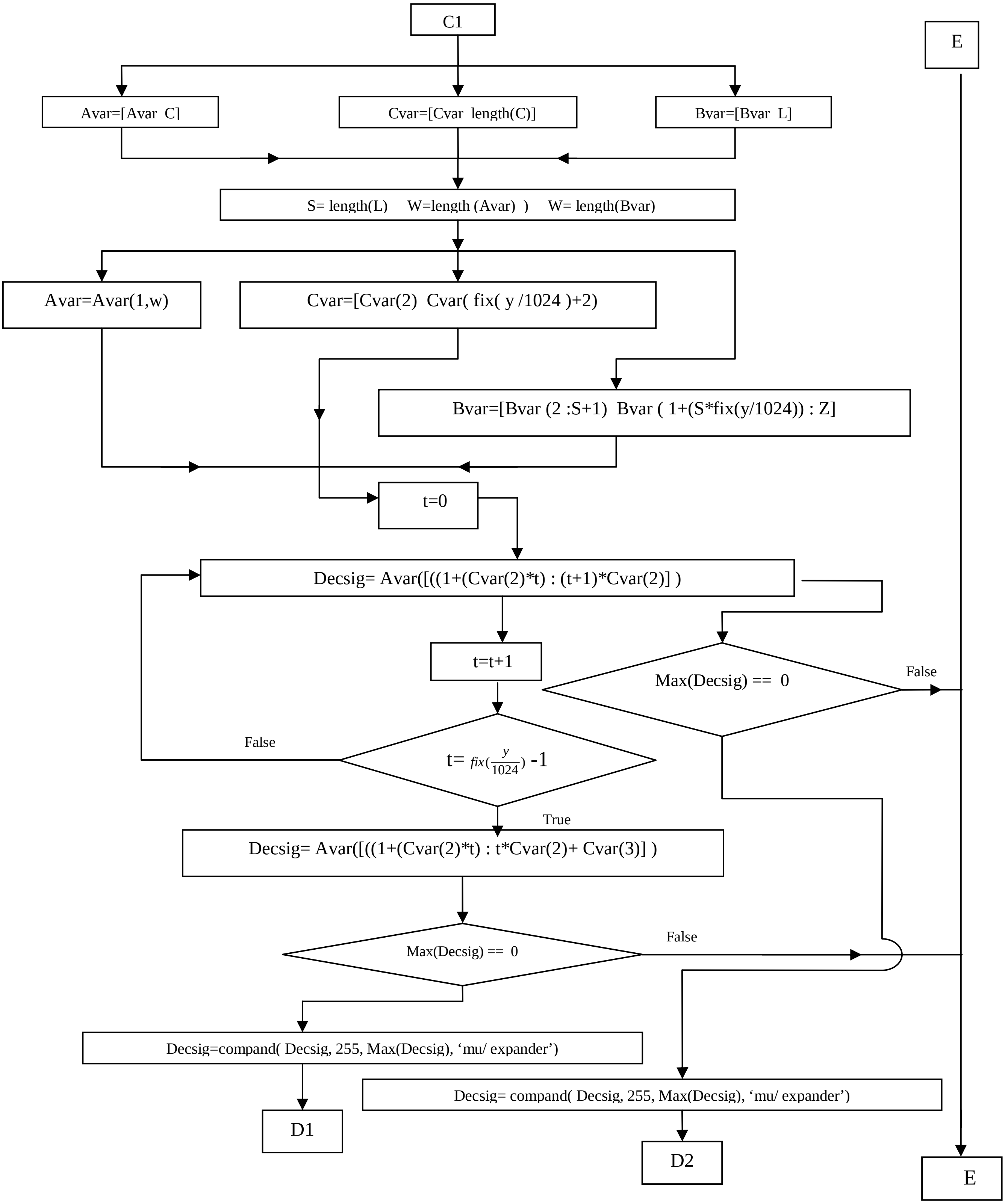


Figure 11. Diagram of the wavelet encoder and decoder (part 4)

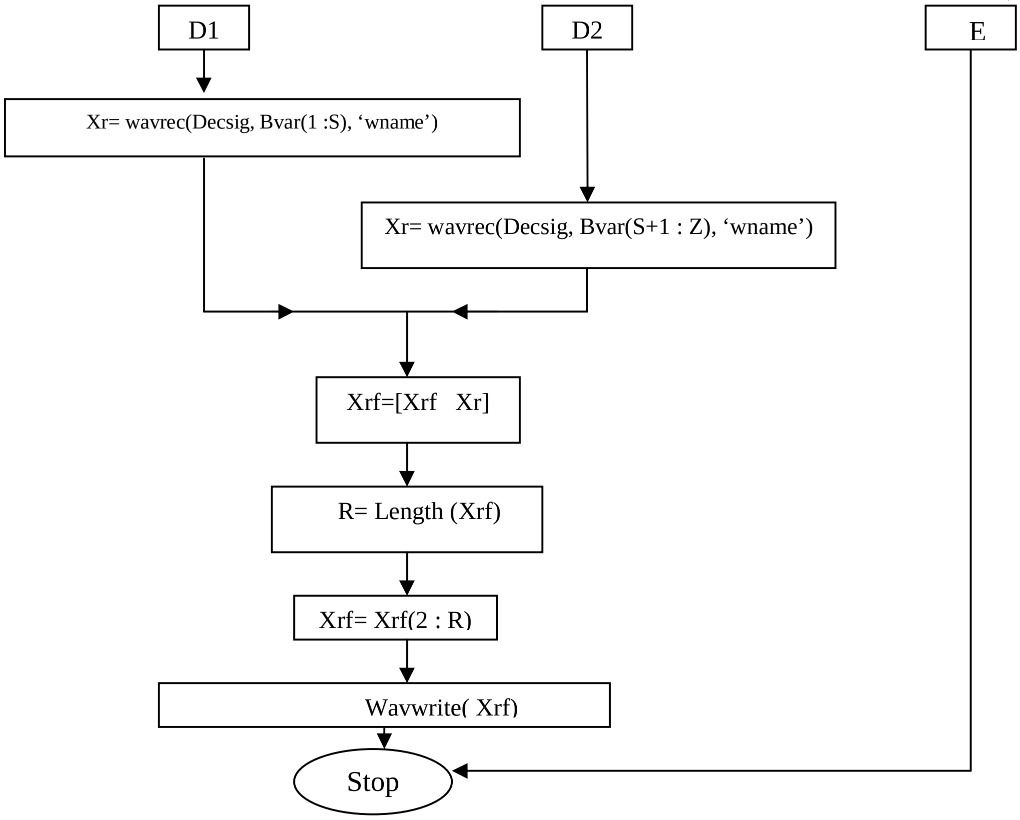


Figure 12. Diagram of the wavelet encoder and decoder (part 5)

VI. Implementation and Results

The proposed wavelet−packet audio codec is realized as m files and simulated using MATLAB software. We adjust parameters such as structure of the decomposition tree, frame size, number of wavelet coefficients, etc. The suitable set of parameters is selected to optimize among decoded audio quality, encoded bit rate and computation complexity.
A number of quantitative parameters can be used to evaluate the performance of the proposed audio wavelet codec, in terms of reconstructed signal quality after decoding. The used quantitative parameters are the Signal to Noise Ratio (SNR) and the compression ratio wich are calculated for different types of wavelet

A. *The signal to noise ratio*

$$SNR = 10.\log_{10}\left(\frac{\sigma_x^2}{\sigma_e^2}\right) \qquad (22)$$

σ_x^2 is the mean square of the speech signal and σ_e^2 is the mean square difference between the original and reconstructed signals.

B. *Compression ratio*

The compression ratio is defined as the quotion between the original audio size file and the compressed one.

$$CR = \left(\frac{Size\_(Original\_File)}{Size\_(Compressed\_File)}\right) \qquad (23)$$

C. *Results*

In order to evaluate the proposed codec, we used for various wavelet ('haar', 'coif', 'morl', 'meyr', 'dB') some types of sound such as Soul, Slow, and Rock. The evaluation is based on the SNR and the compression ratio.

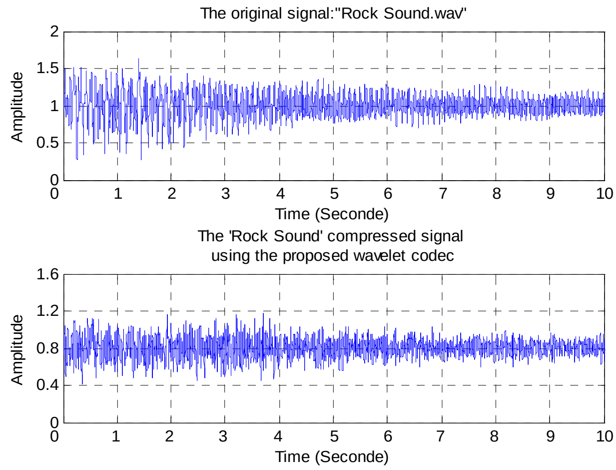


Figure 13. The original signal and the wavelet compressed signal using (bitrate=128Kbits/s wavelet= 'db' 'Rock sound.wav')

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

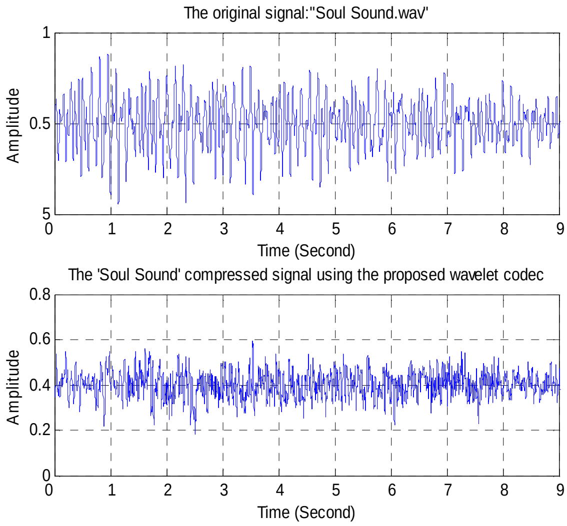


Figure 14. The original signal and the wavelet compressed signal using (bitrate=128Kbits/s wavelet= 'db' 'soul sound.wav')

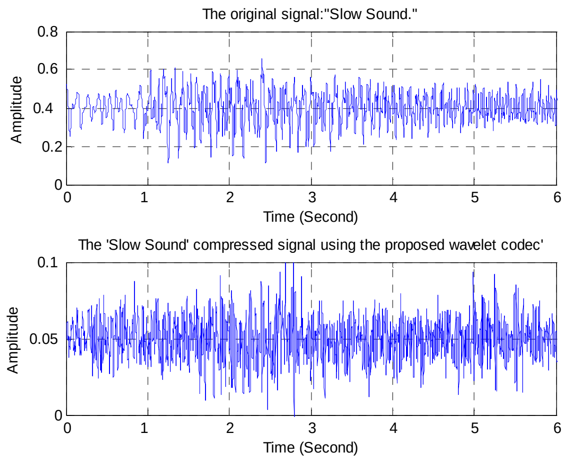


Figure 15. The original signal and the wavelet compressed signal using (bitrate=128Kbits/s wavelet= 'db' 'slow sound.wav')

TABLE I. EVALUATION OF THE PROPOSED CODEC USING DIFFERENT WAVELET (SOUL MUSIC.WAV)

Wavelet name	haar	coif	morl	meyr	db
SNR	30.514	31.379	30.282	30.461	31.012
CR	5.762	6.342	6.117	5.451	7.249

TABLE II. EVALUATION OF THE PROPOSED CODEC USING DIFFERENT WAVELET (SLOW MUSIC.WAV)

Wavelet name	haar	coif	morl	meyr	db
SNR	30.631	30.233	31.681	30.298	30.767
CR	5.663	6.817	6.219	5.358	7.471

TABLE III. EVALUATION OF THE PROPOSED CODEC USING DIFFERENT WAVELET (ROCK MUSIC. WAV)

Wavelet name	haar	coif	morl	meyr	db
SNR	30.091	30.411	31.903	30.104	31.515
CR	5.918	6.628	6.992	5.758	7.735

VII. AUDIO QUALITY MEASURE USING MEAN OPINION SCORE

It is hard to objectively measure the performance of audio compression in the realm of perceptual media, due to the variation in human senses, and the qualitative nature of such a process. However, some attempt has been made to do this. As a measure of quality, the most popular subjective assessment method is the mean opinion scoring where subjects classify the quality of coders on an N-point quality scale. The final result of such tests is an averaged judgement called the *mean opinion score (MOS)*. 5-point adjectival grading scales are in use, one for signal *quality*, and the other one for signal *impairment*, and an associated numbering. The 5-point ITU-R impairment scale of Table 4 is extremely useful if coders with only small impairments have to be graded.

For this purpose, we invited several subjects to hear some wavelet compressed files resulting from the proposed codec based on wavelet analysis. The protocol of evaluation consists in listening to the wavelet compressed sound file. Then, the listeners can listen to it as long as they wish. The listeners are 12: 6 men and 6 women between 15 and 30 years old. Our aim is to determine the best wavelet compression sound quality for each type of sound in a statistic card as shown in Figure 16.

Note:

The sound quality histogram amplitude represent the sum of the integers scores given by the 12 listeners.

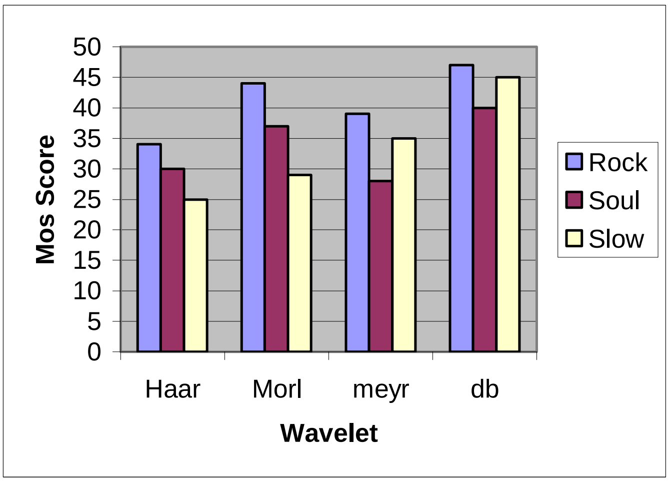


Figure 16. The MOS diagram wavelet listening test

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

TABLE IV. 5-POINTS MOS IMPAIRMENT SCALE

Mean opinion score	Impairment scale
5	Perceptible
4	Perceptible, but not annoying
3	Slightly annoying
2	Annoying
1	Very annoying

VIII. CONCLUSION AND FUTURE WORK

Audio compression coding is currently an active topic for research in the areas of circuit technologies and Digital Signal Processing (DSP). The Wavelet Transform performs very well in the compression of recorded audio signals. Point of view compression ratio, using wavelets can be easily varied, while most other compression techniques have fixed compression ratios.

Further data compaction is possible by exploiting the redundancy in the encoded transform coefficients. A bit encoding scheme could be used to represent the data more efficiently. A common loss-less coding technique is Entropy coding. Two common entropy coding schemes are Prefix coding and tree-structured Huffman coding.

REFERENCES

[1] ISO/IEC 11172-3, "Information technology—coding of moving picture and associated audio for digital storage media at up to about 1.5 Mbits—part 3: audio," 1993.

[2] Z. Hajayej, Etude, mise en oeuvre et évaluation des techniques de paramétrisation perceptive des signaux de parole. Application à la reconnaissance de la parole par les modèles de Markov cachés, PhD Thesis on Electrical Engineering, National Engineering School of Tunis, October 2009

[3] T. Painter and A. Spanias, "Perceptual coding of digital audio," *Proceedings of the IEEE*, vol. 88, no. 4, pp. 451–512, 2000.

[4] M. D. Swanson, B. Zhu, A. H. Tewfik, and L. Boney, "Robust audio watermarking using perceptual masking," *Signal Processing*, vol. 66, no. 3, pp. 337–355, 1998.

[5] P.R. Deshmukh, Multi-wavelet Decomposition for Audio Compression, *IE(I) Journal –ET*, Vol 87, July 2006

[6] Q. Liu, "Digital audio watermarking utilizing discrete wavelet packet transform," M.S. thesis, Institute of Networking and Communication, Chaoyang University of Technology, Taichung, Taiwan, 2004.

[7] J. D. Johnston, "Transform coding of audio signals using perceptual noise criteria," *IEEE Journal on Selected Areas in Communications*, vol. 6, no. 2, pp. 314–323, 1988.

[8] M. Bosi and R. E. Goldberg, *Introduction to Digital Audio Coding and Standards*, Kluwer Academic Publishers, New York, NY, USA, 2003.

[9] D. Sinha and A. H. Tewfik, "Low bit rate transparent audio compression using adapted wavelets," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3463–3479, 1993.

[10] M. R. Zurera, F. L. Ferreras, M. P. J. Amores, S.M. Basc´ on, and N. R. Reyes, "A new algorithm for translating psycho-acoustic information to the wavelet domain," *Signal Processing*, vol. 81, no. 3, pp. 519–531, 2001.

[11] B. Carnero and A. Drygajlo, "Perceptual speech coding and enhancement using frame-synchronized fast wavelet packet transform algorithms," *IEEE Transactions on Signal Processing*, vol. 47, no. 6, pp. 1622–1635, 1999.

[12] C. Wang, Y. C. Tong, An improved critical-band transform processor for speech applications, *Circuits and Systems*, May 2004, pp 461–464

[13] I. Daubechies, Ten Lectures on Wavelets, vol. 61 of CBMSNSF Regional Conference Series in AppliedMathematics, SIAM, Philadelphia, Pa, USA, 1992.

[14] P. Rajmic, J. Vlach, Real-time Audio Processing Via Segmented wavelet Transform, *10th International Conference on Digital Audio Effect* , Bordeaux, France, Sept. 2007

[15] B. Lincoln, "An experimental high fidelity perceptual audio coder," *Project in MUS420 Win97*, March 1998.

AUTHORS PROFILE

K. Abid received the B.S. degree in Electrical Engineering from the National School of Engineering of Tunis, (ENIT), Tunisia, in 2005, and the M.S degree in Automatic and Signal Processing in 2006 from the same school. He started preparing his Ph.D. degree in Electrical Engineering in 2007. His research interesets in Audio Compression Using Multiresolution Analysis

K. Ouni received the M.Sc. from Ecole Nationale d'Ingénieurs de Sfax in 1998, the Ph.D. from Ecole Nationale d'Ingénieurs de Tunis, (ENIT), in 2003, and the HDR in 2007 from the same institute. He has published more than 70 papers in Journals and Proceedings. Professor Kaïs Ouni is currently the Electrical Engineering Department Head at Institut Supérieur des Technologies Médicales de Tunis (ISTMT), Tunisia. He is also a researcher at Systems and Signal Processing Laboratory (LSTS), ENIT, Tunisia. His researches concern speech and biomedical signal processing. He is Member of the Acoustical Society of America and ISCA (International Speech Communication Association).

N. Ell ouze was born in 19 December, 1945. He received a Ph.D. degree in 1977 at INP (Toulouse- France), and Electronic Engineering Diploma from ENSEEIHT in 1968 University P. Sabatier. in 1978. Pr. Ellouze joined the Electrical Engineering Department at ENIT (Tunisia). In 1990, he became Professor in signal processing, digital signal processing and stochastic process. He was the head of the Electrical Department from 1978 to 1983 and General Manager and President of IRSIT from 1987-1994. He is now Director of Research

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

(IJCSIS) International Journal of Computer Science and Information Security,
Vol. 08, No. 04, July 2010

A Tropos Based Requirement Engineering Frameworks for Self Adaptive Systems

Farah Noman
Department of Computer Science
National University of Computer and Emerging Sciences
Karachi, Pakistan
farah.nomansaghir@gmail.com

Zafar Nasir
Department of Computer Science
National University of Computer and Emerging Sciences
Karachi, Pakistan
zafar.nasir@nu.edu.pk

***Abstract*—The applications developed during the current era are deployed in environments which change over the course of time. These changes if occur in a normal application would require re-work so that design and architectural level updates should be implemented to cater to the newly changed application environment. This is turns results in wasted effort and increased cost for the system maintenance. Hence there arise a need for systems that are able to alter their functioning so as to adapt to the changing environmental needs and to heal themselves automatically from likely errors and system failures without the need of human intervention. Such systems are known as Self Healing, Self Adaptive or Autonomic Systems.**

The approaches and frameworks used for gathering, analyzing and specifying requirements for Self Adaptive system are quite different from the traditional life cycle and require a different line of action from the processes which are followed when capturing requirements for a normal system whose environments are relatively stable and all the system states are known before hand.

This research focuses on analyzing the various methods, techniques and frameworks for gathering requirements for Self Adaptive systems. A Tropos based approach has also been proposed for requirements engineering for self adaptive system.

Keyword-component; Self-adaptive Systems; Requirement Engineering; Autonomic Systems; Agent oriented Methodologies

I. INTRODUCTION

The changing needs of the modern era have led forward to an application domain where the underlying environment of the system often changes. Such changes occur as part of the normal working course of the system and are considered as part of the common working environment. Thus if such changes are catered to so as to ensure the smooth functioning of the systems, the system should also be designed in a way so that effective decision making can be performed by the system to alter and adapt their behaviours to the changing environmental needs [1].

Although most of the self healing systems development approaches are in infancy but many industry leaders such as Microsoft, IBM and SUN and performing extensive research for the development of autonomic systems. Thus such systems are meant to extend their capabilities to restore to a normal functioning state from an error state. Thus if such an ability is incorporated in the current systems then unlimited environmental changes can be effectively catered and the systems can be made available for infinite periods thus reducing the system maintenance cost and in turn saving organizational budgets [2] [3].

The paper presents a Tropos based framework which focuses on the requirements engineering perspective for self adaptive systems. The Tropos methodology is incorporated with the RELAX language specification constructs to cater all the uncertainty factors ingrained in the self adaptive systems execution environment.

II. RELATED WORK / LITERATURE REVIEW

During the recent years the area of software engineering have witnessed tremendous amount of research and analysis in the field of self adaptive systems which is conducted in almost all the verticals of software development of self adaptive system right from their inception to modelling to engineering and quality assurance [5]. This has resulted in the onset of a vast number of techniques for each of the system development domain. A brief discussion of these techniques is as follows:

The core concepts of most of the requirement engineering techniques for self adaptive systems are derived from the concept of goal models where a requirements model is developed through the usage of specific environmental conditions and goal states. Agents are allocated to goal traversal graphs where alternative conditions are to be considered or negotiated with respect to the specific system variables [6]. One of the earliest goal modelling notations such as KAOS (Knowledge Acquisition in Automated Specification) [7] and i* [8] presents a comprehensive overview of goal models however in such approaches there is no scope for catering to uncertainty or adaptively of the system. Hence researchers have proposed extensions of the goal directed requirement acquisition techniques for self adaptive systems.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Amongst the many extended approaches, one of the prominent one is that of the KAOS specification language incorporated with the "Adapt-Operator" [9]. Another approach focuses on formal methods such as the RELAX specification language for converting traditional requirements statements to the ones that are able to cater to the uncertainty of the self managing systems [10] [11]. Extension to this approach provides an amalgamated view of the goal graphs and RELAX by placing specific language constructs in goal graphs nodes where adaptation is specifically required [12].

An extended approach defines the pathway of transforming the requirement specification to the software architecture. This approach derives the self adaptive component model from the KAOS goal model [13]. Another similar approach focuses on the development of a formalized notation in terms of Unified Modelling Language (UML) for the requirements specification of self adaptive systems [14].

A different framework focuses on the classification of goals into soft goals and hard goals thus defining a road map for converting the goal models to architectural level diagrams by providing the concept of Autonomic Elements (AEs) [15]. Also focusing on the system architecture, a research paper proposes a high level approach to facilitate the smooth transition from the system requirements specifications to the system architecture [16].

Utilizing the strengths of the i* modelling language for representing goals models, a different approach focuses on the requirements specifications for embedded self adaptive systems [17]. Focused on the distributed software systems, one more approach defines a framework to develop self adaptive systems by adopting a Belief-Desire-Intention (BDI) [4]. Extension to the Tropos methodology is provided in this approach by modelling environmental variables, goal precedence and priority and goal correlations for catering to the flexibility required in all self managing systems.

A similar approach specifically focus on modelling the requirements specifications for self adaptive systems by using the advanced version of TROPOS and BDI agents i.e. Tropos4AS [18]. A related technique illustrates a tool (TAOM4E) that based on the Tropos4AS [19]. An additional Tropos related approach has been proposed for requirements specification of self managing systems that provides a framework to translate traditionally captured requirements to adaptive requirements through the combination of Tropos goal oriented methodology and the domain [20].

III. A Proposed Requirements Specification Framework for Self Adaptive System

A. Framework Introduction

The study of the various techniques proposed by various researchers have led to the conclusion that all the proposed requirement specification and engineering techniques proposed for self adaptive systems lack some aspect or the other. Some techniques make use of the goal models by defining system alternatives and environmental constraints while completely ignoring the system environmental uncertainty factors whilst the other which caters to this uncertainty factors does not provide a systematic mechanism for transforming the developed specifications into formal architectures, and design.

Although some of the work has been performed in defining an amalgamated view of the goals models with formal specifications language for self adaptive systems [12] however, they take into consideration the general goal models without focusing on specific ones which can expedite the requirements engineering process and which are flexible enough to cater to the requirements change during the system execution environment.

Thus we propose a consolidated framework for requirements engineering for self adaptive system by incorporating the TROPOS requirements specification methodology with the RELAX specifications to cater to the environmental uncertainty factors and to reap the benefits of TROPOS for easy requirements modelling and catering to the early and late stage of the requirements engineering life cycle.

The subsequent section briefly presents the main rationale behind the proposition of the requirements specification framework. The next sections explain the TROPOS requirements engineering framework followed by the introduction of the RELAX specifications. The next section will shed light on the actual process and various steps of the extended proposed framework followed by a case study that illustrates the practical implementation of the framework.

B. Rationale for the Proposition of the Extended Modelling Framework

It is a well argued fact that specifying the requirements for self managing system is a testing task due to the level of uncertainty in the underlying system environment where the values of the various quantities are unpredictable beforehand. Also the techniques for requirements specification be it a goal modelling framework or a formal specification language, they fail to cover all the aspects of catering to this flexibility in the requirements set. An idea has been proposed as a prospective future work for catering to this problem which suggest combining the strengths of the RELAX specification language with the contemporary goal models for proposing a process model through which thorough requirements specification can be performed by the analysts [11]. Thus based on the idea we propose a process model for specifying the requirements for self managing system by combing the strengths of the Tropos Modelling Framework and the RELAX Specification language so that a robust requirements set can be developed.

The main advantages of this approach are multi-fold. The Tropos Modelling approach presents a well known and stable

process for requirements specification right from the inception phase of the Early Requirements gathering to the system Architectural design and Implementation. On the other hand the RELAX specification presents specific language constructs as well as formal notations and semantics descriptions for incorporating flexibility in the requirements set for self adaptive systems.

However, if Tropos alone is applied to gathering requirements for self adaptive systems, it cannot fulfil the prerequisite for an adaptive set where specific environmental variables can be quantified so that partial fulfilment of goals can be taken as satisfactory for iterating a goal path. The RELAX specification alone lacks a formal process of gathering the early requirements and then iteratively working upon that set to transform the requirements to the architectural design and the implemented system.

Also there are a number of agent oriented methodologies presented in literature for system modelling. However their application and span of usage is restricted to their level of maturity. Studying a number of research literature related to the comparison of agent oriented methodologies [21] [22], it is also concluded that the Tropos is one single methodology that is not only stable and relatively mature but it also lays a strong focus on the early requirements analysis where the domain stakeholders and their intentions are identified so that the main reason for the system development can be acknowledged at the initial stages.

Hence this proposed framework presents a process of identifying the initial set of the requirements and working upon them for refining them to cater to the underlying necessities for self adaptive systems. A case study is also presented for applying the framework for modelling requirements of a self adaptive system such as SmartHome for proving the practicality of the approach.

C. Requirements Engineering in Tropos

The Tropos is an agent oriented software development methodology which is developed by a group of authors from various universities in Canada and Italy. The Tropos is a unique agent oriented software engineering methodology that drives its strengths from three basic factors such as it keenly focuses on the activities that precede the prospective requirement engineering practice by collecting information about how and why the intended system is envisioned to achieve the organizational goals. Hence firstly, it focuses on the broad picture of the system so that a perspective is established and ground is set for detailed requirements analysis [23].

Secondly, it provides means of specifying and designing the system right from the requirements analysis to the system architecture, design and implementation phases in a systematic manner. It drives its strength from the notions of actors, hard goals, soft goals, plans, resources, mean-end links and intentional dependencies.

Lastly, the methodology is built on the novel idea that the system to be should be modelled in an incremental fashion, refining it at every subsequent stage from the conceptual level to the executable artefacts through the usage of a sequence of transformational steps. Hence the major advantage of this methodology over the other modelling notations is derived from the fact that it not only concentrates of the what or how perspectives of the system but also on the why aspect of the system [24] [25] [26].

The Tropos model is extensively inspired by the Eric and Yu's framework for requirements engineering which presents the idea of actor, goals and actor dependency links as primitive concepts [27]. The Tropos model propose a goal based requirements modelling technique to capture and analyse both the business and system requirements, risk and security and location requirements [26].

1) *Phases of Tropos:* The Tropos methodology is divided into the following four major phases of the software development lifecycle [28][29][30]:

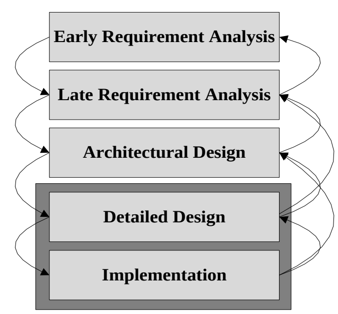


Fig. 1. Tropos Iterative and Incremental Approach

D. RELAX: A Specification Language

RELAX is a requirements specification language developed explicitly for self adaptive system by the researchers at the universities of England, USA and France. The language constructs explicitly support the expression of the environmental uncertainty in the requirements [11]. This environmental uncertainty is the core of the self adaptive systems where flexible requirements needs to be specified so that changes to the environment and other conditions should not affect the normal functioning of the system and the system should gracefully alter its execution path to cater to such changes.

The RELAX vocabulary is specified in a set of its operators which are organized into modal, temporal and ordinal operators and uncertainty factors [11]. The following table describes the RELAX vocabulary constructs.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

TABLE I
RELAX VOCABULARY AND OPERATORS

RELAX OPERATOR	DESCRIPTION
Modal Operators	
SHALL	a requirement must hold
MAY…OR	a requirement specifies one or more alternatives
Temporal Operators	
EVENTUALLY	a requirement must hold eventually
UNTIL	a requirement must hold until a future position
BEFORE, AFTER	a requirement must hold before or after a particular event
IN	a requirement must hold during a particular time interval
AS EARLY, LATE AS POSSIBLE	a requirement specifies something that should hold as soon as possible or should be delayed as long
AS CLOSE AS POSSIBLE TO [frequency]	a requirement specifies something that happens repeatedly but the frequency may be relaxed
Ordinal Operators	
AS CLOSE AS POSSIBLE TO [quantity]	a requirement specifies a countable quantity but the exact count may be relaxed
AS MANY, FEW AS POSSIBLE	a requirement specifies a countable quantity but the exact count may be relaxed
Uncertainty Factors	
ENV	defines a set of properties that define the system's environment
MON	defines a set of properties that can be monitored by the system
REL	defines the relationship between the ENV and MON properties
DEP	identifies the dependencies between the (relaxed and invariant) requirements

E. The Proposed Requirements Modelling Approach

The distinguishing characteristic of the Self Managing system is that there can a number of pathways for realizing a high level system objective and a set of runtime environmental quantities and variables dictate which particular pathway realization is appropriate at a particular time. Hence for incorporating such a kind of variation in the execution paths at runtime, the goal modelling approach for requirements specification offers support to identify and visualize various different alternatives for satisfying the overall objectives and goals of the system [31]. These alternatives for a system requirement can be due to differences in the system non-functional goals such as performance, reliability, robustness etc or due to the ambiguity in the system environment which can affect the goal paths that are iterated at run time. Hence the goal modelling approach proves to be a fine approach for modelling goal decompositions in terms of its subsequent low level goals.

Hence the modelling approach is primarily based on the Tropos Requirements modelling framework that work on the Eric Yu's i* modelling methodology which is one the renowned framework for requirements specifications using goal models [29]. The Tropos presents additional benefits in terms of early requirements gathering which adds extra flexibility and span of usage to our proposed framework.

Adjoining the Tropos goal models is the RELAX specification which will be incorporated at the locations where there remains scope for the system uncertainty and where requirements can be temporarily relaxed to support the system adaptation process. Thus if there exist some goal paths where non-critical requirements can be partially neglected in order to satisfy other short term critical requirements then RELAX specification presents specific vocabulary and constructs that can effectively cater to such necessity.

F. The Proposed Methodology Process Model

The steps for developing the requirements specification using the mentioned modelling approach consists of a step wise procedure for systematically performing the system analysis so that general as well as adaptive requirements can be effectively gathered. The Fig. 2 describes the high level flow diagram of the modelling steps. The detailed are described as follows:

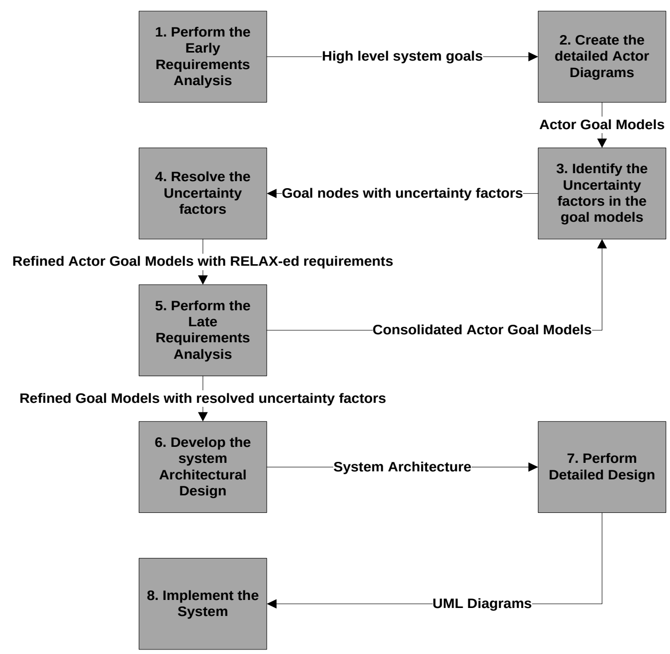


Fig. 2. The Proposed Process Model

1) Step 1. Perform the Early Requirements Analysis: Identify and analyse the system stakeholders and their intentions for the usage of the system. Enlist the set of actors and their respective high level system goals that they need to achieve using the system. Create the high level Actor diagrams in this step. This is performed in congruence with the general approach adopted for the Tropos Modelling framework.

2) Step 2. Create the detailed Actor Diagrams: Perform a thorough analysis on each of the high level goal identified in

the Step 1 for decomposing the goal into subsequent low-level goals unless a leaf node is achieved which is one single goal, resource or plan which cannot be further decomposed. This step is also performed following the methodology presented by the Tropos framework.

3) *Step 3. Identify the Uncertainty factors in the goal models:* Iterate the goal graph developed in the previous step using a bottom up approach from the leaf nodes to the top nodes for the identification of the uncertainty factors that can pose hindrance in the achievement of the preceding top level goals. These factors can be the varying environmental conditions that should be monitored for smooth functioning of the system.

4) *Step 4. Resolve the Uncertainty factors:* The goal graph nodes that are marked with the uncertainty factors needs to be thoroughly studied for mitigating and resolving the identified factors. There can be following approaches for the resolution of the identified uncertainty factors:

4.1) *Do not perform any refinement:* If the identified environmental uncertainty factors does not threaten the satisfaction of a low level goal achievement in relation to its preceding higher level goal hence the node should be left as it is with no refinement iteration.

4.2) *Refine the leaf node with further sublevels:* Sometimes the goal uncertainty factors can be resolved by performing analysis on the leaf goal node and further breaking it up to low level goals so that the factors can be effectively captured, measured and analysed during the process of the goal graph execution.

4.3) *Introduce the RELAX operators:* As discussed earlier, the uncertainty factors that are part of the system execution environment are sometimes of a nature that only their partial fulfilments can prove to be good enough for the complete fulfilment of the high level goals. Hence for marking such state and conditions, the RELAX operators needs to be introduced in the goal nodes for providing the specific uncertainty factors and RELAX operators for precisely measuring the flexibility in the environmental quantities.

4.4) *Create a new high level Actor diagram:* It is also an observed scenario that at time the effect of the environmental uncertainty factors is so intense that no goal refinement or relaxation can help in following the normal goal graph or the execution path of the system. Hence for such environmental conditions, new high level goals needs to be identified for the system actors which will be executed according to the new set of environmental conditions. Also this should be noted that this is the most expensive form of resolution since it will require the reapplication of the Steps 1-4 to this newly created goal.

5) *Step 5. Perform the Late Requirements Analysis:* The late requirements analysis needs to be performed where the target system is added as a new actor in the goal diagram together with its functions and quantities. This requirements analysis models the new target system actor and its social dependencies on the other actors. This in turn refines the goal model by placing the concept of the overall system and its interaction scenarios with all the system actors.

6) *Step 6. Develop the system Architectural Design:* The system architectural design focuses on the system's global architecture in terms of the sub-systems (actors) interconnected through data and control flows (dependencies). The architecture is articulated in a three step fashion where the overall architecture in terms of extended actor diagrams is performed and then the capabilities to be performed by the actor dependencies are defined followed by defining a set of agent types with one or more different capabilities (agent assignment). This step is performed according to the traditional Tropos Modelling framework.

7) *Step 7. Perform Detailed Design:* The detailed system design is related to the system agent's micro level activities such creation of the system capability diagrams which is part of the family of the UML Activity diagrams and the Agent interaction diagrams. This step completely models the system in terms of the UML diagrams to aid in the system implementation phase.

8) *Step 8. Implement the System:* The implementation is self explanatory where the actual system development is performed against the detailed design implemented in the previous steps.

IV. Application of the Requirements Specification Framework – A Case Study

For validating the practical implementation of the proposed requirements specification framework, we are presenting a case study of a SmartHome application used for assisted living. The step by step implementation of such a system and its requirements gathering and specification using the proposed modelling framework is described below:

1) *Step 1. Perform the Early Requirements Analysis* During this early requirements analysis phase where the intentions of the system stakeholders are identified and analysed and are modelled into the subsequent goal diagrams. The goal diagrams are modelled in terms of the hard goals, soft goals, plans and resources as per the convention of the Tropos Modelling framework. The high level actor diagram for the SmartHome system is depicted in Fig 3.

2) *Step 2. Create the detailed Actor Diagrams:* The detailed Actor diagrams are an extension of the strategic or social dependency model created in Step 1. This step includes the extension of each and every goal defined in the previous step through analysis so that dependencies with other actors are

refined. The top level goal which is under consideration is AND-OR decomposed into sub-goals (such as hard goals and soft goals), plans and resources. For each of the leaf node of the goal graph means tasks are identified which can be further AND-OR decomposed. Additionally, the needed resources are also established.

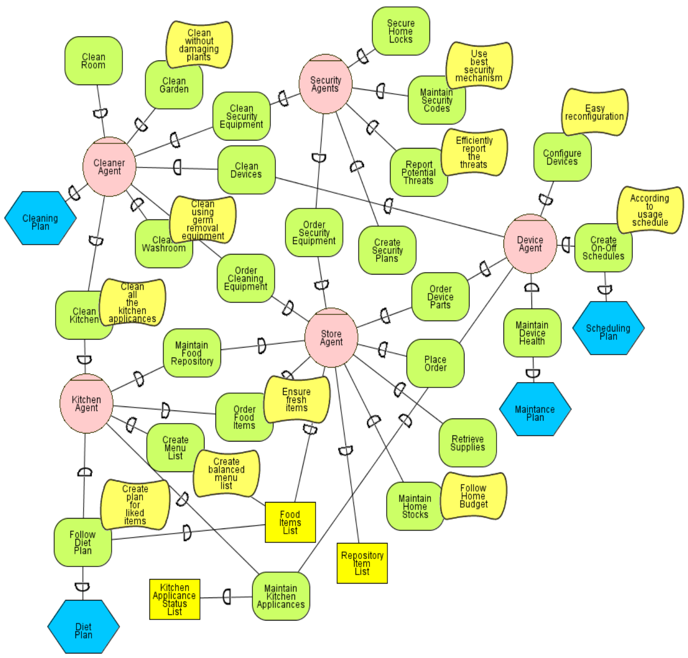


Fig. 3. Actor Diagram or the Social Dependency model

For the sample case study for the SmartHome, we have chosen the Cleaner Agent's "Clean Security Equipment" goal for refinement for creating the detailed Actor Goal Model. This is depicted in Fig 4

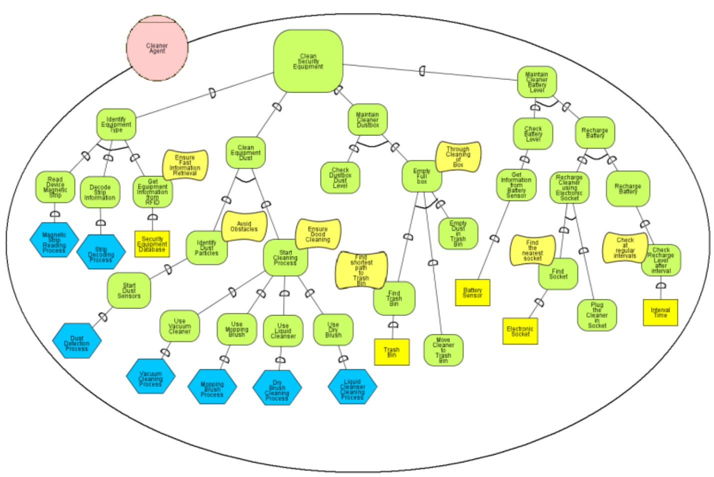


Fig. 4. Cleaner Agent Goal Model for "Clean Security Equipment"

3) Step 3. Identify the Uncertainty factors in the goal models:

The step in which the uncertainty factors of the environment are identified to introduce flexibility in the goal nodes. In the "Clean Security Equipment" goal graph some goal nodes have been identified that can become victim of uncertainty factors hence the RELAX specification operators needs to be identified for specifying conditions for their partial fulfilment. The marked goal nodes are shown in Fig 5.

For example, the goal node marked as "Recharge Battery" has an uncertainty factors related to the level of its battery recharge. The partial satisfaction of the goal such as a particular threshold level of the battery recharge levels can prove to be enough for the working of the Cleaner Agent.

Step 4. Resolve the Uncertainty factors: Hence the "Recharge Battery" can be relaxed by introducing a RELAX operator known as "AS CLOSE AS POSSIBLE TO FULL". Thus more RELAX operators are introduced at the "Start Dust Sensors", "Start Cleaning Process" and "Find Trash Bin" goal nodes. The Goal graph with RELAX-ed nodes is shown in Fig 6.

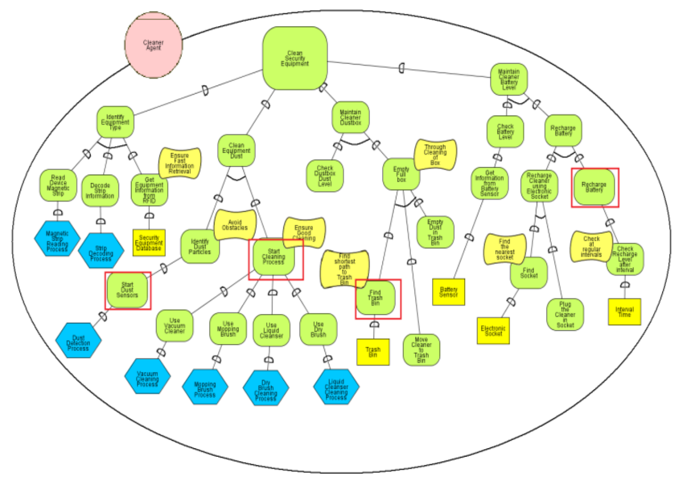


Fig. 5. Goal Graph with marked Uncertainty factors

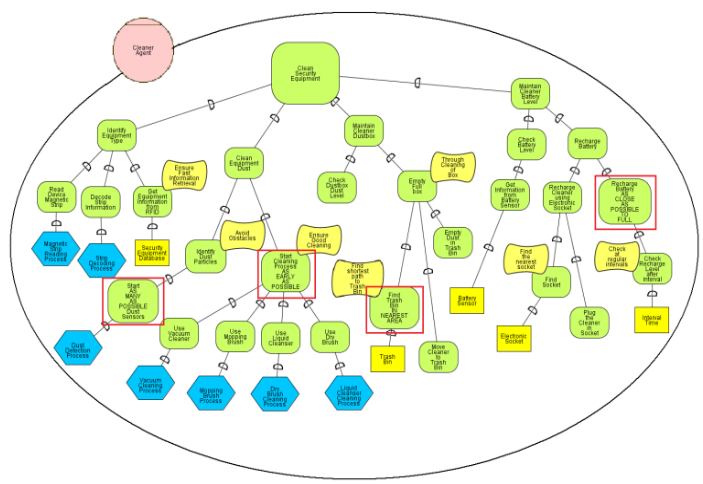


Fig. 6. Goal graph with RELAX-ed nodes

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

4) *Step 5. Perform the Late Requirements Analysis:* The phase of late requirements analysis work upon introducing the SmartHome as actor named as "SmartHome Agent" where in similar manner the actor goals and interdependencies with the other system actors such as goal graphs of all the goals for each of identified system actor such as "Cleaner Agent", "Security Agent", "Kitchen Agent", "Store Agent" and "Device Agent" are modelled. A subset of the overall actor diagram for the System Agent is depicted in Fig. 7.

5) *Step 6. Develop the system Architectural Design*: The architectural design of the system works upon decomposing and refining the system actors diagrams identified during the late requirements analysis phase by describing the structure of the overall architecture pattern together with further identifying the interactions and dependencies between different actors by considering them as agents with the perspective of the overall system.

Various kinds of diagrams are modelled during this phase such as the system overview diagrams, architectural style diagrams and the overall system decomposition diagrams. Also the system capabilities definition as well as the agent definition in terms of the defined capabilities is also developed during this phase. A partial capabilities definition table for the Cleaner Agent's Clean Security Equipment goal are described in Table II.

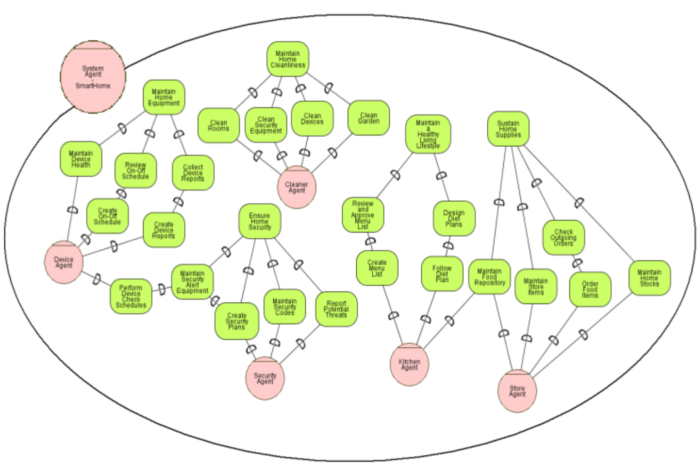


Fig. 7. Late Requirement Analysis Goal graph

6) *Step 7. Perform Detailed Design:* The detailed design phase focus on making the system UML diagrams for the system. Various kinds of UML diagrams are developed during this phase so that all the aspects of the system such as it transitions, states, capabilities and attributes can be thoroughly identified.

A SmartHome component diagram is shown in Fig 8.

TABLE II
SYSTEM CAPABILITIES DEFINITION TABLE

Agent Name	Capability ID	Means-End
Cleaner Agent	CP-001	Identify Equipment Type; Read Device Magnetic Strip
	CP-002	Identify Equipment Type; Decode Strip Information
	CP-003	Identify Equipment Type; Get Equipment Information from RFID
	CP-004	Identify Dust Particles; Start AS MANY AS POSSIBLE Dust Sensors
	CP-005	Start Cleaning Process AS EARLY AS POSSIBLE; Use Vacuum Cleaner
	CP-006	Start Cleaning Process AS EARLY AS POSSIBLE; Use Mopping Brush
	CP-007	Start Cleaning Process AS EARLY AS POSSIBLE; Use Liquid Cleanser
	CP-008	Start Cleaning Process AS EARLY AS POSSIBLE; Use Dry Brush
		…
		…
	CP-(N-3)	Clean Security Equipment; Identify Equipment Type
	CP-(N-2)	Clean Security Equipment; Clean Equipment Dust
	CP-(N-1)	Clean Security Equipment; Maintain Cleaner Dust box
	CP-(N)	Clean Security Equipment; Maintain Cleaner Batter Level

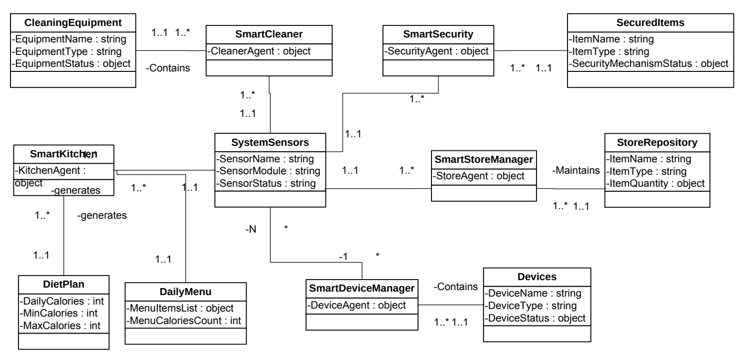


Fig. 8. SmartHome Component Diagram

7) *Step 8. Implement the System:* Various tools are available using which the UML diagrams can be converted into the code skeletons and eventually the system actual implementation code. Making use of the same tools, the development code of the SmartHome application will be developed.

V. CONCLUSION AND FUTURE WORK

The requirements engineering activity serves as the basis for the development and analysis of any system. However, the system inherent behaviour and its underlying environment have a huge effect on the practices followed for requirements engineering. From the span of a large number of requirement

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

engineering frameworks, goal models are usually adopted for specifying requirements for self adaptive systems which have a number of environmental uncertainty factors that put risk to the accomplishment of various system goals and tasks. Thus in this paper we presented a comprehensive view of the various requirements engineering frameworks and practices following for requirement elicitation, specification, verification and monitoring of self adaptive systems.

Additionally, we presented a comprehensive set of the quality requirements which can serve as basis for the evaluation of a newly proposed framework. These quality requirements measure the effectiveness of a proposed technique against a variety of factors such the level of the extension of the technique to various domains, the number of uncertainty factors it caters to and the extent of the support of the technique to the various phases of the system development life cycle.

Lastly based on the mentioned quality requirements, we proposed a goal based modelling approach for the specification of self managing systems where various environmental uncertainty factors pose a threat to the smooth execution of the system goal paths. The proposed framework takes its concepts from the well known Tropos Modelling framework and the RELAX specification language by amalgamating the concepts of both for creating a framework that fulfils almost all of the quality requirements of the requirement specification for the self adaptive system.

A great number of future directions sprint up after the proposition of the respective framework where the development of a formal tool for modelling such methodology tops the list. Other future directions include the development of a formal language which can translate the late requirements analysis goal models into the written requirements specification so as to aid the system analysts. Various language modelling techniques are already present in the literature and their adaptation and extension in relation to the proposed framework can serve as a promising future direction.

REFERENCES

[1] Stephan Weibelzahl: “Problems and Pitfalls in Evaluating Adaptive Systems”

[2] Debanjan Ghosh, Raj Sharman , H. Raghav Rao and Shambhu Upadhyaya, “Self-healing systems - survey and synthesis”

[3] Yuriy Brun, Giovanna Di Marzo Serugendo, Cristina Gacek, Holger Giese, Holger Kienle, Marin Litoiu, Hausi M¨uller, Mauro Pezz`e, and Mary Shaw: “Engineering Self-Adaptive Systems through Feedback Loops”

[4] Mirko Morandini, Loris Penserini, Anna Perini: “Towards Goal-Oriented Development of Self-Adaptive Systems”

[5] B. H. C. Cheng, H. Giese, P. Inverardi, J. Magee, and R. de Lemos: “Software engineering for self-adaptive systems: A research road map, Dagstuhl-seminar on software engineering for self-adaptive systems”

[6] Axel van Lamsweerde: “Goal-Oriented Requirements Engineering: A Guided Tour”

[7] Anne Dardenne, Axel van Lamsweerde and Stephen Fickas: “Goal-directed Requirements Acquisition”

[8] Yu, E.S.K.: “Towards modeling and reasoning support for early-phase requirements engineering”

[9] Greg Brown, Betty H.C. Cheng, Heather Goldsby, Ji Zhang: “Goal Oriented Specification of Adaptation Requirements Engineering in Adaptive Systems”

[10] Jon Whittle, Pete Sawyer, Nelly Bencomo, Betty H.C. Cheng: “A Language for Self-Adaptive System Requirements”

[11] Jon Whittle, Pete Sawyer, Nelly Bencomo, Betty H.C. Chengy and Jean-Michel Bruelz: “RELAX: Incorporating Uncertainty into the Specification of Self-Adaptive Systems”

[12] Betty H.C. Cheng, Pete Sawyer, Nelly Bencomo, Jon Whittle: “A Goal-Based Modeling Approach to Develop Requirements of an Adaptive System with Environmental Uncertainty”

[13] Shan Tang, Xin Peng, Yijun Yu, and Wenyun Zhao: “Goal-Directed Modeling of Self-adaptive Software Architecture”

[14] Yijun Yu, Alexei Lapouchnian, Sotirios Liaskos, John Mylopoulos, Julio C.S.P. Leite: “From Goals to High-Variability Software Design”

[15] Alexei Lapouchnian, Sotirios Liaskos, John Mylopoulos and Yijun Yu: “Towards Requirements-Driven Autonomic Systems”

[16] Matthew J. Hawthorne and Dewayne E. Perry: “Exploiting Architectural Prescriptions for Self-Managing, Self-Adaptive Systems: A Position Paper”

[17] Pete Sawyer, Nelly Bencomo, Danny Hughes, Paul Grace, Heather J. Goldsby, Betty H. C. Cheng: “Visualizing the Analysis of Dynamically Adaptive Systems Using i* and DSLs”

[18] Mirko Morandini, Loris Penserini, Anna Perini: “Modelling Self-Adaptivity: A Goal-Oriented Approach”

[19] Mirko Morandini, Loris Penserini, Anna Perini: “Automated Mapping from Goal Models to Self-Adaptive Systems”

[20] Nauman A. Qureshi and Anna Perini: “Engineering Adaptive Requirements”

[21] Khanh Hoa Dam and Michael Winikoff: “Comparing AgentOriented Methodologies”

[22] Paolo Giorgini, Anna Perini, John Mylopoulos, Fausto Giunchiglia and Paolo Brescian: “Agent-Oriented Software Development: A Case Study”

[23] P. Bresciani, P. Giorgini, F. Giunchiglia, J. Mylopoulos, and A. Perini: “Tropos: An agent-oriented software development methodology”

[24] Paolo Bresciani and Fabrizio Sannicol: “Requirements Analysis in Tropos:a self referencing example”

[25] P. Bresciani, A. Perini, P. Giorgini, F. Giunchiglia, and J. Mylopoulos: “Modelling Early Requirements in Tropos: A Transformation Based Approach”

[26] http://www.troposproject.org: “The Tropos Project”

[27] E. Yu: “Modelling Strategic Relationships for Process Reengineering”

[28] John Mylopoulos and Jaelson Castro: “Tropos: A Framework for Requirements-Driven Software Development”

[29] Paolo Giorgini, John Mylopoulos and Roberto Sebastiani: “Goal-Oriented Requirements Analysis and Reasoning in the Tropos Methodology”

[30] Maddalena Garzetti, Paolo Giorgini, John Mylopoulos, and Fabrizio Sannicol-o: “Applying Tropos Methodology to a real case study: Complexity and Criticality Analysis”

[31] Axel van Lamsweerde and Emmanuel Letier: “Handling Obstacles in Goal-Oriented Requirements Engineering”

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Fuzzy Logic in a Low Speed Cruise-Controlled Automobile

Mary Lourde R., Waris Sami Misbah,
Department of Electrical & Electronics Engineering
BITS, Pilani-Dubai, Dubai International Academic City, U.A.E

Abstract — **Traffic congestion is a major problem that drivers face these days Long rush hours exhibit both mental and physical toll on a driver. This paper describes the design of cruise control system based on fuzzy logic, in order to reduce the workload on a driver during traffic congestion. The proposed low speed cruise control system operates by sensing the speed and headway distance of the preceding vehicle and controlling the host vehicle's speed accordingly. The vehicle speed is controlled by controlling throttle and the brakes. The fuzzy logic based cruise controlled vehicle is simulated using MATLAB Simulink and the results are presented in this paper.**

Keywords - fuzzy logic, cruise control, low speed, and traffic congestion.

I. INTRODUCTION

A cruise control system is a general feature found in most of the automobiles today. A basic cruise-controlled car travels at constant speed set by the driver, allowing automatic movement of the vehicle without the driver pressing the accelerator. The driver sets the speed as desired and then the cruise control system maintains that speed by controlling the throttle of the car. A typical cruise control system comes with features such as acceleration, coasting and resume functions.

Since the cruise control system replaces the driver, it must be able to imitate human behavior. The use of fuzzy logic is an ideal tool for this purpose. Fuzzy logic, which also means imprecise logic, when applied to system makes it user friendly. A fuzzy system involves a set of linguistic rules applied on set of input and output parameters, in order to control a system.

Conventional cruise control systems generally operate at speeds greater than 40 km/h; mostly used by drivers at highways. For speed lower than this, the vehicle needs to be controlled manually. A cruise control system that operates at lower speed is rarely available.

II. SOFTWARE USED TO SIMULATE THE SYSTEM

The software used for the modeling of the system is MATLAB/SIMULINK. It has several toolboxes available such as the Fuzzy Logic Toolbox, SIMULINK, Image processing, Simdriveline, SimMechanics, SimScape etc. all of which can be integrated with a control system. This allows the user to develop most of the real world conditions in MATLAB. It is the most commonly used platform by most of the scientific organizations.

In order to make the fuzzy logic cruise control system more realistic, we need to model a commercially available car on MATLAB and then integrate the fuzzy cruise control system to the vehicle model. The car chosen for modeling is Toyota Yaris 2007 Sedan[1]. The reason behind the selection of this car is the availability of technical information of its control system.

III. MODELING OF THE VEHICLE ON MATLAB

The modeling of the vehicle's drive train and dynamics is done by mapping on the specifications of a Toyota Yaris onto a demonstration model in MATLAB[6]. The automatic drivetrain model available in SIMULINK is taken as the base model for the system development.

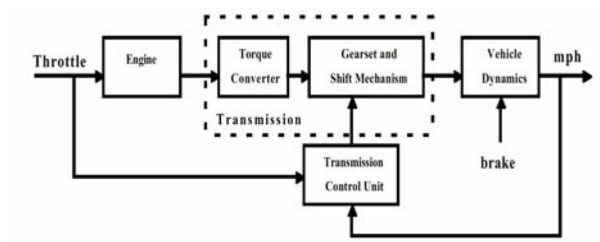


Figure 1. Block diagram of basic Drivetrain system

The inputs to the drive train are the throttle opening and brake torque. The engine, vehicle dynamics and the automatic transmission have been modeled using non-linear differential equations. The transmission control unit has been modeled in STATEFLOW as it involves decision-making based on the current state of the vehicle.

Reference [6] gives the complete details of modeling equations used by Mathworks to design this drive train. The complete Drivetrain model used for simulation is shown in figure 3 below. The engine subsystem was modified according to the engine-torque curve of Toyota Yaris. Default transmission gear ratios were modeled to that of Toyota Yaris. The shift schedule used by the shift logic block and the vehicle dynamics parameters were modified by the corresponding Toyota Yaris values. Any subsystem for which the corresponding Toyota Yaris specification was not available, the default value used by Mathworks has been retained.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

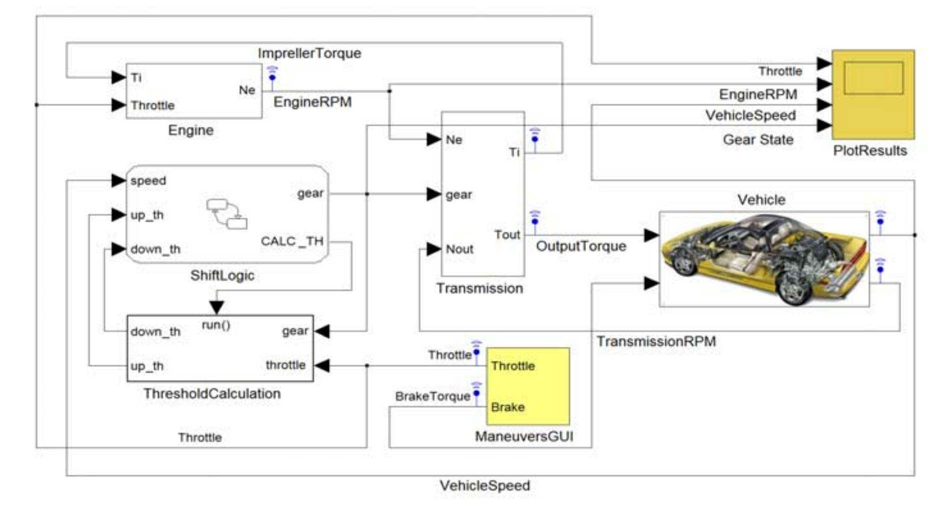


Figure 2. Complete Drivetrain SIMULINK model used for the simulation [6]

Simulation One

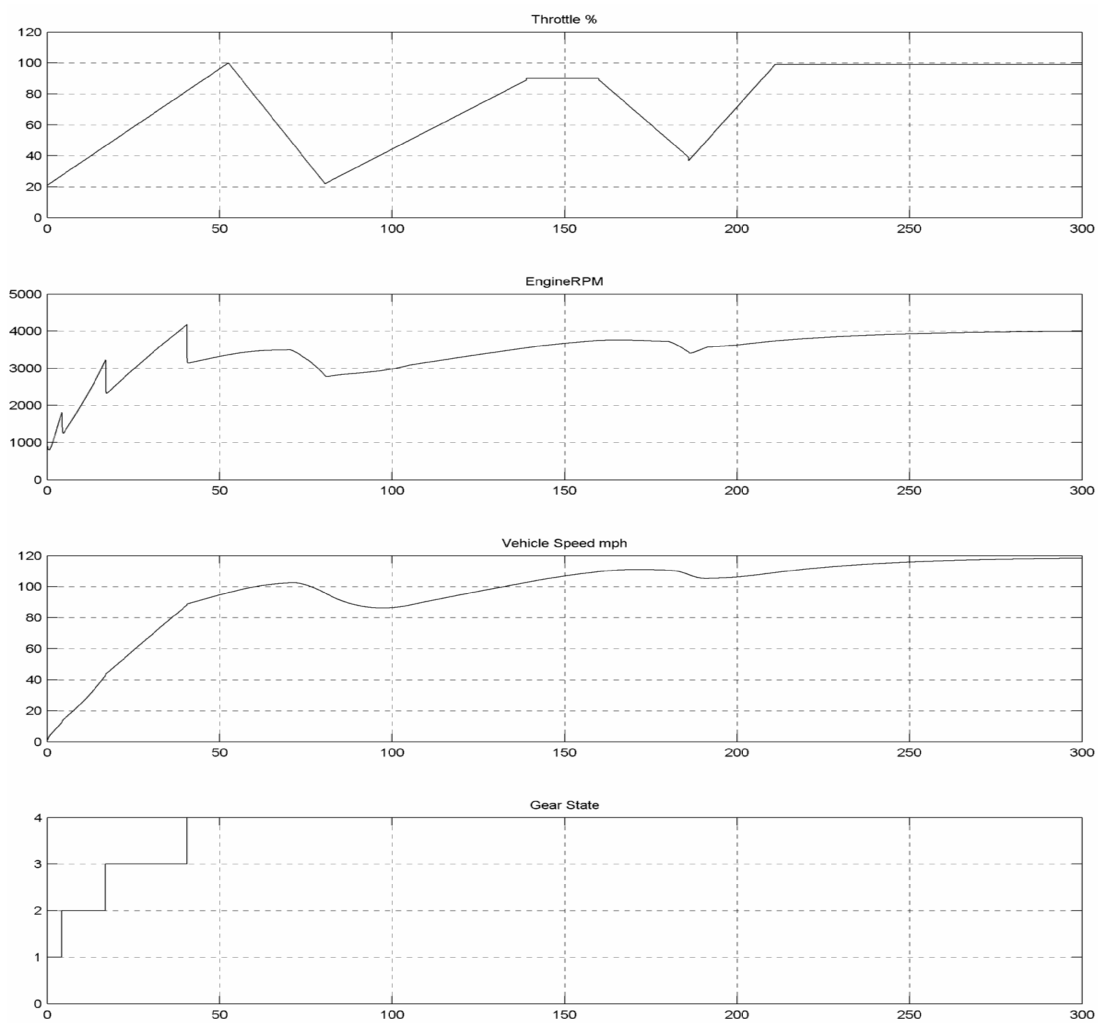


Figure 3. Simulation 1 results of a Toyota Yaris 2007 automobile under normal running environment

IV. MODEL TESTING OF THE AUTOMOBILE

The vehicle model is tested for accuracy and compared with the actual operation of Toyota Yaris. Several simulations, with various inputs, were carried out to verify the simulation model of the vehicle of which two set of results are given below.

The vehicle's throttle profile are selected to simulate the real time operation of the engine and vehicle speeds and the gear positions are also observed to verify the working of the vehicle. From the simulation1 results shown in figure 3 above it is seen that the vehicle speed as well as the engine speed closely follows as required by the instruction specified by the throttle profile of the vehicle.

As we can see from the graph, the vehicle speed reaches its maximum value of 120 mph, which is the top speed of Toyota Yaris. The maximum engine rpm reached is 4100, which is about 100 revolutions less than that of Toyota Yaris. The up shifts take place at 10 mph and 40 mph, which are close to the shift schedule of Toyota Yaris.

Simulation two

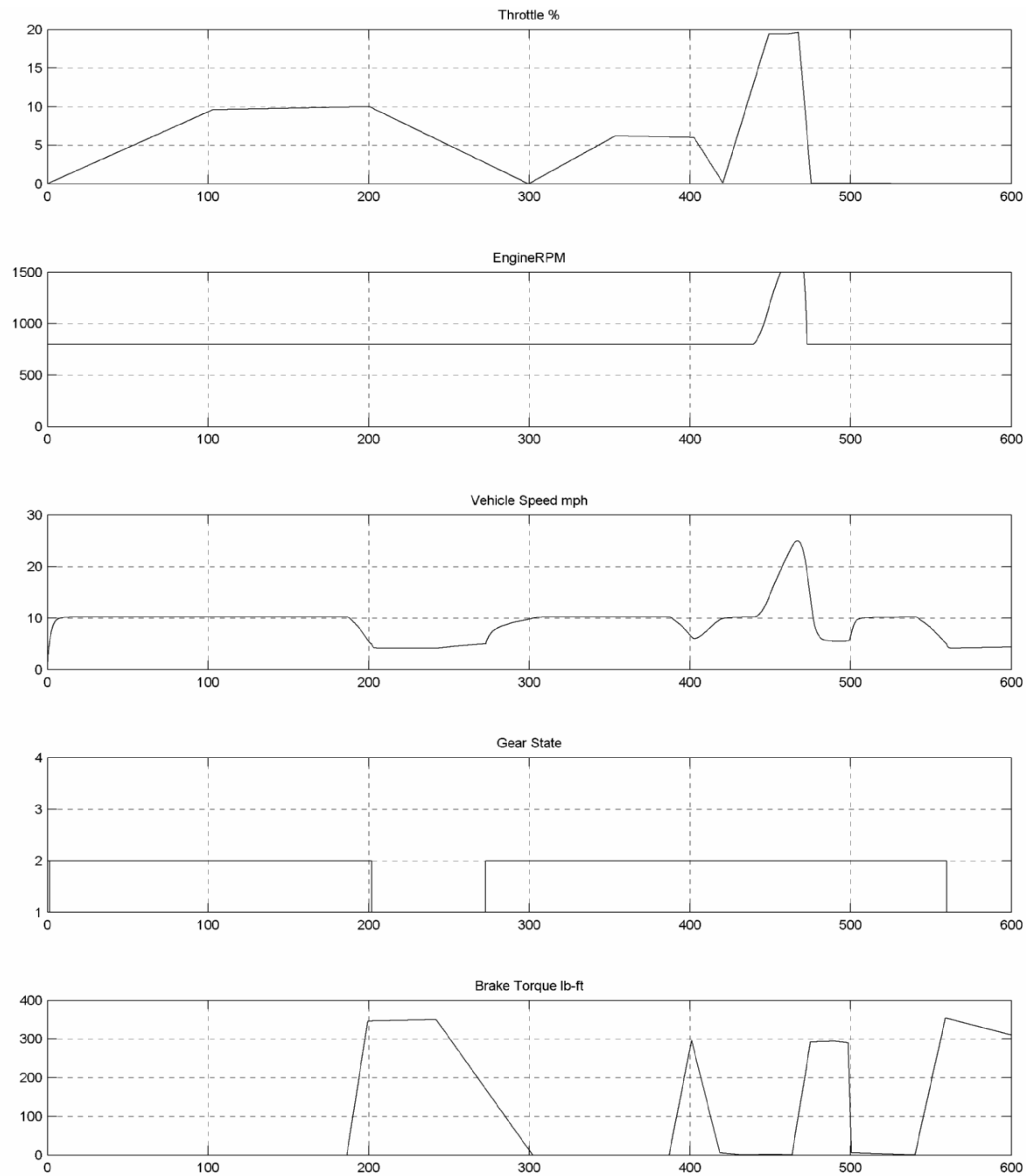


Figure 4. Simulation 2 results of the automobile in heavy traffic environment

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Figure 4 shows a typical example of a heavy traffic environment as the vehicle stays below 10 mph for a major amount of time during the simulation. The simulation1 is carried out for 300 seconds and the simulation2 is done for 600 secs. In simulation 2, in addition to the throttle profile, the second input brake torque is also applied. These two simulation results proves the model of the automobile in the Simulink platform.

V. DESIGN OF THE FUZZY LOGIC CRUISE CONTROL SYSTEM

The objective of the fuzzy controller is to control the vehicle in a congested slow moving traffic environment. Therefore, the design of the fuzzy controller must be based on the variables that affect the vehicle's movement in such an environment

A. Input to the Fuzzy Controller

In order to control the vehicle longitudinally, a sensor must detect a vehicle ahead and provide the distance and relative velocity of that vehicle with respect to the host vehicle. Therefore, a 24 GHz radar sensor can be installed in the vehicle since it has a very short range [7].

The sensor is not modeled in MATLAB/SIMULINK instead, an output profile coming from a sensor is assumed and provided as an input to the fuzzy controller. There are two output profiles from the sensor. They are relative velocity and relative distance between two vehicles. The figure below shows two typical unfiltered measurement profiles from a sensor [8].

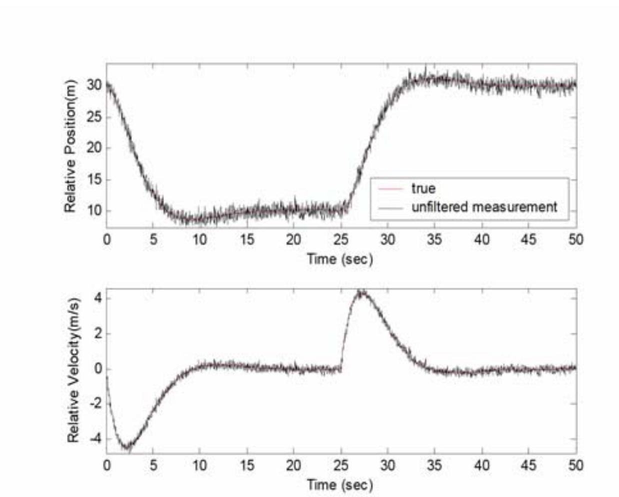


Figure 5. Output profiles from a radar sensor

It can be observed from the above profiles, that when relative velocity is positive, the distance between the vehicles increases. The distance between cars will keep on increasing, as long the relative velocity is positive. The rate of change of distance varies the rate of change of relative velocity. When relative velocity becomes negative, the distance between the vehicles starts decreasing. Relative distance decreases as long as the relative velocity remains negative.

B. Definition of Input Membership Functions

Inputs to the fuzzy controller are relative velocity and the distance between the two vehicles. Therefore, we need to define membership functions for fuzzy variables relative velocity and relative distance. The range for the fuzzy variable relative velocity chosen is -10 km/h to +10 km/h. The speed range chosen is the typical range that vehicles travel in a congested traffic situation. The range of relative distance is chosen 0.5 to 2m, which again is the typical range in a congested traffic situation. Most of the membership functions chosen in this system are triangular shaped as it has less parameters and responds rapidly when compared to other functions. This helps the inference system, to make decisions more effectively when compared to other membership functions. Literature also shows that most commonly used membership function is triangular shaped, due to its effectiveness in a real-time environment and economic feasibility.

1) Relative Velocity

The relative velocity fuzzy variable has the following linguistic values: mildly negative (MN), mildly positive (MP), negative (N), positive (P), very negative (VN), very positive (VP) and null (N). The figure 6 below shows the membership function definition for relative velocity.

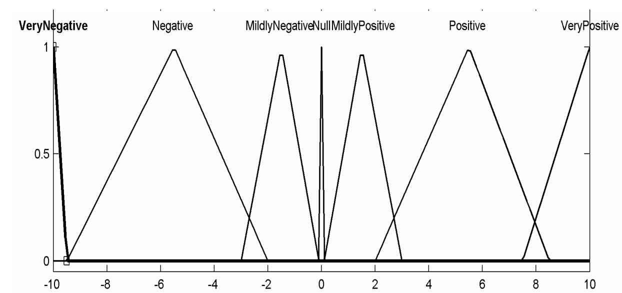


Figure 6. Membership function for Relative Velocity

2) Relative Distance

The fuzzy variable for relative distance has the following linguistic values: very very close (VVC), very close (VC), close (C), Mildly Close (MC), distant (D), very distant (VD) and very very distant (VVD). The figure 7 below shows the membership definition for relative distance

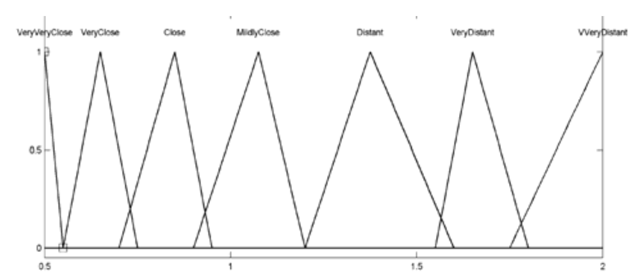


Figure 7. Membership function for relative distance (0.5 - 2 m).

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

C. Output of the Fuzzy Controller

1) Throttle Control

The linguistic values for throttle control are down (D), down very low (DVL), down low (DL) down medium (DM), up (U), up very low (UVL) up low (UL), up medium (UM) as shown in figure 8 below. The throttle control membership functions are classified into two major categories, Up and Down. These two are further divided into subcategories normal, low, very low and medium. This is to enhance the control of the vehicle at different relative velocities and distances.

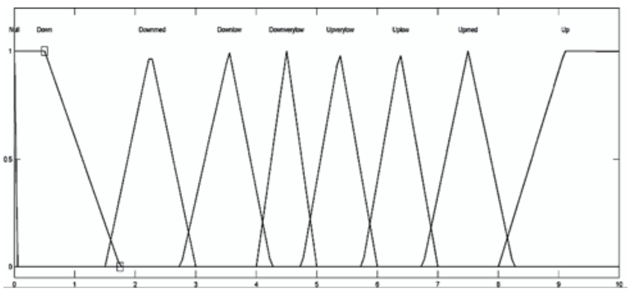


Figure 8. Membership function for Throttle Control

Majority of the membership functions cross each other. The membership functions have been made to cross in order to allow for smooth transition from one membership function to other, while the input values are changing. If the cross points are not included, discontinuities might arise in the operation of the controller as no rule will be fired at the end of each membership function.

2) Brake Control

Controlling the brake along with the throttle allows better control of the car. Since the vehicle is assumed to travel in congested traffic situation, it will be required to stop frequently. The linguistic values for the fuzzy variable brake control are no brake (NB), slow brake (SB), medium brake (MB), high brake (HB) and hard brake (HDB). The figure 9 below shows the membership functions of the brake control

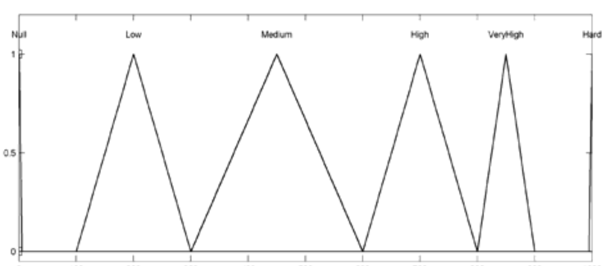


Figure 9. Membership function for brake control

D. Construction of the Rule Base

The construction of rules base determines relation between input and output membership functions. This means that the fuzzy controller will give a certain output depending upon the input and the rules that are executed. A set of rule guides the fuzzy inference system to make decision regarding the control of the output variable. A rule can take three forms conditional, unconditional and an assignment. In this system case rules must be formed in order to control the output variables throttle and brake based on the input variables, relative velocity and relative distance. A set of 31 rules for throttle control and another set of 34 rules for brake control is composed. All the rules are conditional. Table 1 shows set of rules for brake control and Table 2 shows the rule set for throttle control.

TABLE I. RULE BASE FOR BRAKE CONTROL

RD RS	VN	N	MN	NULL	MP	P	VP
VVC	HRD	VH	-	HRD	L	L	-
VC	VH	H	VH	-	L	L	-
C	-	H	M	-	L	L	VL
MC	-	H	L	-	L	VL	-
D	-	M	M	-	L	NULL	NULL
VD	-	M	M	-	L	NULL	NULL
VVD	-	M	M	NULL	-	NULL	NULL

TABLE II. RULE BASE FOR THROTTLE CONTROL

RD RS	VN	N	MN	NULL	MP	P	VP
VVC	D	D	D	NULL	DM	-	-
VC	DM	DM	DM	-	DM	UVL	-
C	-	DM	DL	-	DVL	UVL	-
MC	-	DL	DVL	-	UVL	UL	UM
D	-	DV L	DV L	-	UL	UM	U
VD	-	-	-	-	UM	U	U
VVD	-	-	-	U	-	U	U

E. Defuzzification Method

The final output of the fuzzy controller must be a discrete value. In order to achieve this, a method called defuzzification must be applied. The centroid method was used to defuzzify the output fuzzy variables, since all the membership function definition are triangular. The centroid method makes it easier to defuzzify a triangular shaped membership function and reduces the overall computational task allowing the system to respond effectively

F. Selection of Inference Method

The fuzzy logic toolbox provided in MATLAB has two inference methods, Mamdani inference method and the Sugeno method .The Mamdani method is commonly employed in most of the applications due to the fact that the Mamdani method gives outputs as fuzzy variables. Whereas Sugeno method gives linear output. As both the outputs (throttle and brake) used in this controller are fuzzy variables, Mamdani method is selected for the inference.

G. Fuzzy Logic Cruise Control System model in MATLAB/SIMULINK

The fuzzy logic cruise control system needs to be integrated with the vehicle model. The vehicle model

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

has throttle opening and brake torque as its inputs. Two fuzzy controllers are used to control the throttle and brake respectively. Figure 10 shows the complete MATLAB-SIMULINK block diagram of the Fuzzy Logic Cruise Controlled automobile.

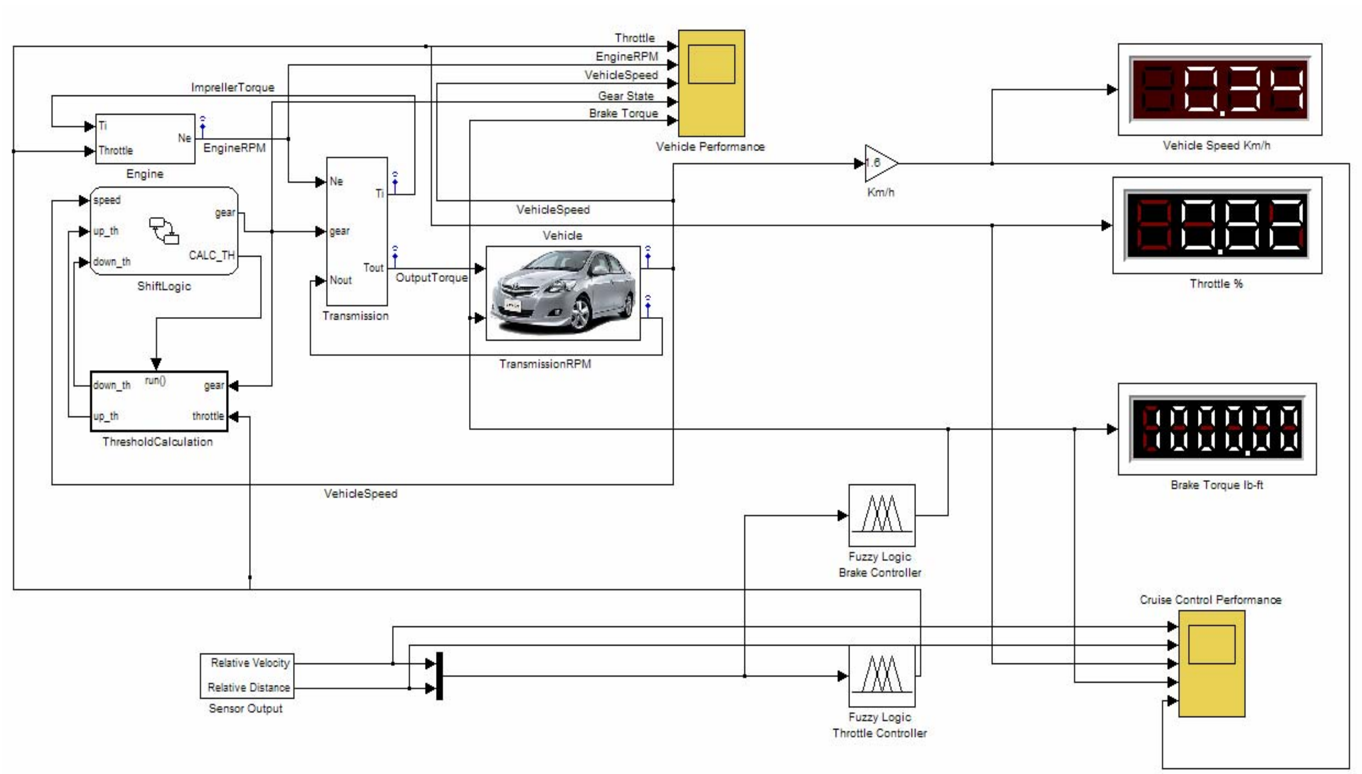


Figure 10. Complete MATLAB/SIMULINK model of the Fuzzy Logic Cruise Control System

VI. SIMULATION OF THE FUZZY LOGIC CRUISE CONTROL SYSTEM

A. Simulation3 - Low speed cruise control system case i

The fuzzy logic cruise control system simulation is simulated for thirty minutes with the assumed profile for relative velocity and relative distance as shown in figure 11. The throttle position, brake torque and vehicle speed are the output variables observed. Initally as the relative velocity and distance increase, the speed of the vehicle increases. An increase in throttle opening and the lowering of the brake can be observed. At 100s, relative velocity begins to decrease, however the relative distance still increases, although at a lower rate. This is due to the fact that relative distance will increase as long as the relative velocity is positive. The relative distance starts to decrease the moment, relative velocity crosses the zero mark. The throttle is reduced and the brake torque applied increases as the relative velocity and relative distance and consequently, we see a decrease in vehicle speed. The vehicle speed continues to decrease further as relative velocity increases in the negative direction.

At 600s, the relative velocity becomes zero and the relative distance reaches its minimum value. Therefore, the fuzzy cruise control system increases the brake torque to 1000 lb-ft and throttle is reduced to about 1%. Both the relative velocity and relative distance remain constant during the next 200s and so does the brake torque and throttle. We can observe that during this period of 200s that the vehicle speed remains zero. Therefore, it can concluded during this period, the vehicles are at a stop. This is a typical situation of vehicles being stuck in a heavy traffic environment. The cruise control system has successfully detected the vehicle approaching the preceding vehicle and brings the vehicle to stop when the minimum distance is reached.

At 800s, as both the relative velocity and distance begin to rise,the throttle increases and the brake torque is lowered allowing the car to acclerate and maintain constant speed for about 200s.Then as relative velocity and distance increase to a higher value, we see a further rise in throttle value and lowering of brake. During the final stage of the simulation, the relative velocity and distance begin to decrease and therefore the cruise control system lowers the throttle and increases the brake in order to adjust the speed of the vehicle. Finally, as the vehicle approaches the preceding vehicle, a drop in the throttle and an increase in the brake torque can be seen, which bring the vehicle to a stop.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

During the simulation, we observe that vehicle speed remains constant for a certain period, even though both the relative velocity and distance are varying. This is because of the fuzzy logic controller being assigned rules to maintain a certain speed of the vehicle for a certain range of relative velocity and relative distance. The fuzzy controller takes into account the degree of change in both the relative velocity and the distance. This condition can be observed in the following simulations as well.

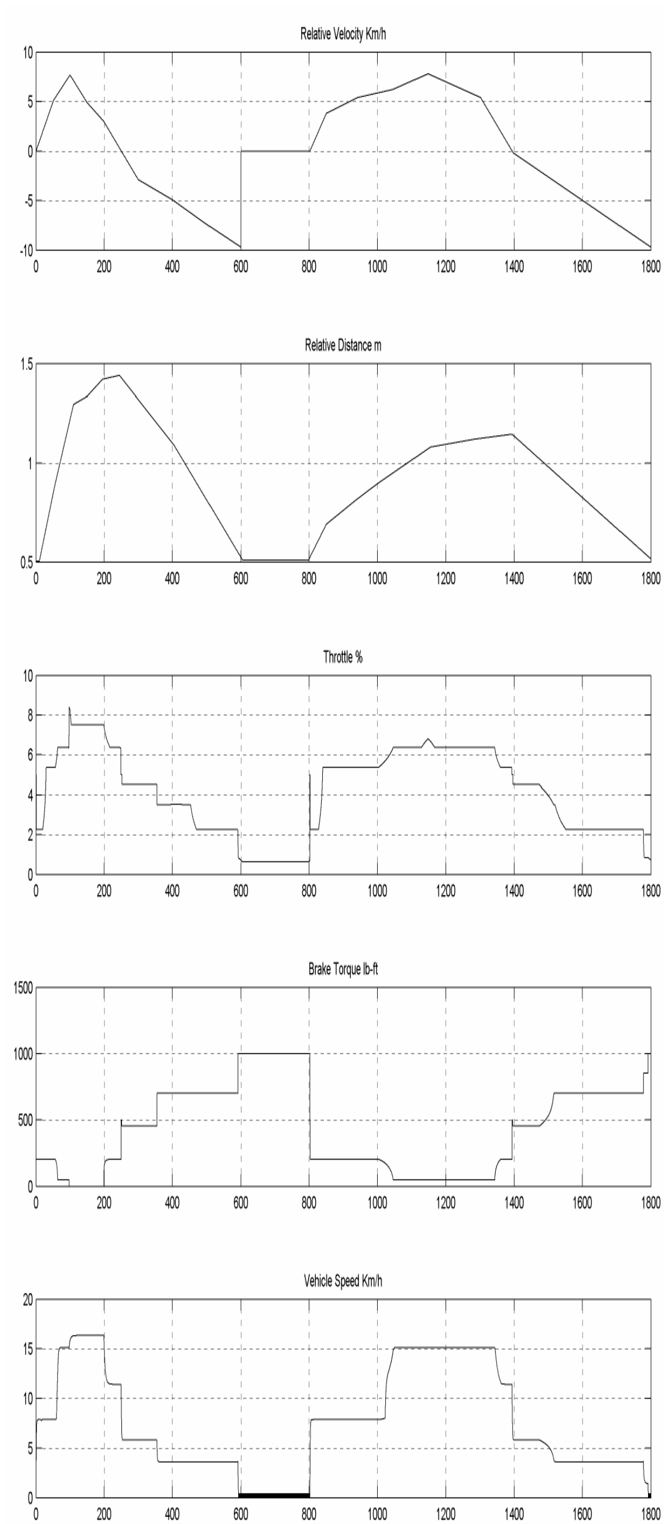


Figure 11. Simulation3- Low speed cruise controlled system case i

B. Simulation4 –Low speed cruise control system case ii

In the intial phase of simulation, the vehicle speed increases as the both realtive velocity and distance increase. A further rise in vehicle speed(16 km/h at about 200s) is observed as shown in the figure 12 when relative velocity and distance increase further. During this period appropriate changes made by the fuzzy logic controller, can be observed, to the throttle and the brake torque to bring about this rise in vehicle speed. The vehicle speed remains constant for about next 600s.

At 400s, the relative velocity reaches zero and the relative distance reaches the maximum value (2m) Both of them remain at that value for the next 200s. During this period, the both the throttle and brake torque remain the same and so does the vehicle speed. This is sitaution, where both the vehicles move at constant speed and headway distance. The fuzzy logic controller has managed to detect this and has responded correctly.

At 800s, the relative velocity decreases negative and subsequently the relative distance alos begins to decrease. Therefore, we see a drop in the throttle and an increase in the brake torque applied. The vehicle speed, thereby decreases to 5.80 km/h and remains at this speed until the the relative velocity crossed the zero. As mentioned in the analysis of the previous simulation, the vehicle speed remains the same even though both relative velocity and distance decrease. The reason being the same as metioned earlier.

In the final phase, the vehicle speed increases as both the relative velocity and distance increase. Appropriate changes in throttle and brake can be observed during this period.

Discontinous spikes in the throttle and brake graphs can be observed whenever the relative velocity crosses the zero the mark. This reason behind this is that no rule is executed by the fuzzy controller at that very instant. However, this does not have an effect on the vehicle speed,since the time span for which these discontuinities occur is negligible.

C. Simulation5 – Low speed cruise control system case iii

As the simulation begins, the vehicle acclerates as both relative velocity and distance increase. Further changes in vehicle speed can be observed as relative velocity and distance continue to increase. The fuzzy logic controller has adjusted to vary the throttle and brake appropriately to raise the vehicle speed. The results are shown in the figure 13.

The vehicle speed begins to decrease at 300s due to the realtive velocity crossing the zero mark. To lower the vehicle the speed the fuzzy controller has increased the brake torque and has simulatanouelsy lowered the throttle. As the relative velocity and distance continue to decrease, we can observe the further decrease of throttle and increase of brake torque in order to lower the vehicle speed.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

The vehicle speed rises again the moment the relative velocity crosses the zero mark at 600s. The fuzzy controller the varies the throttle and the brake torque to bring about this change in vehicle speed.

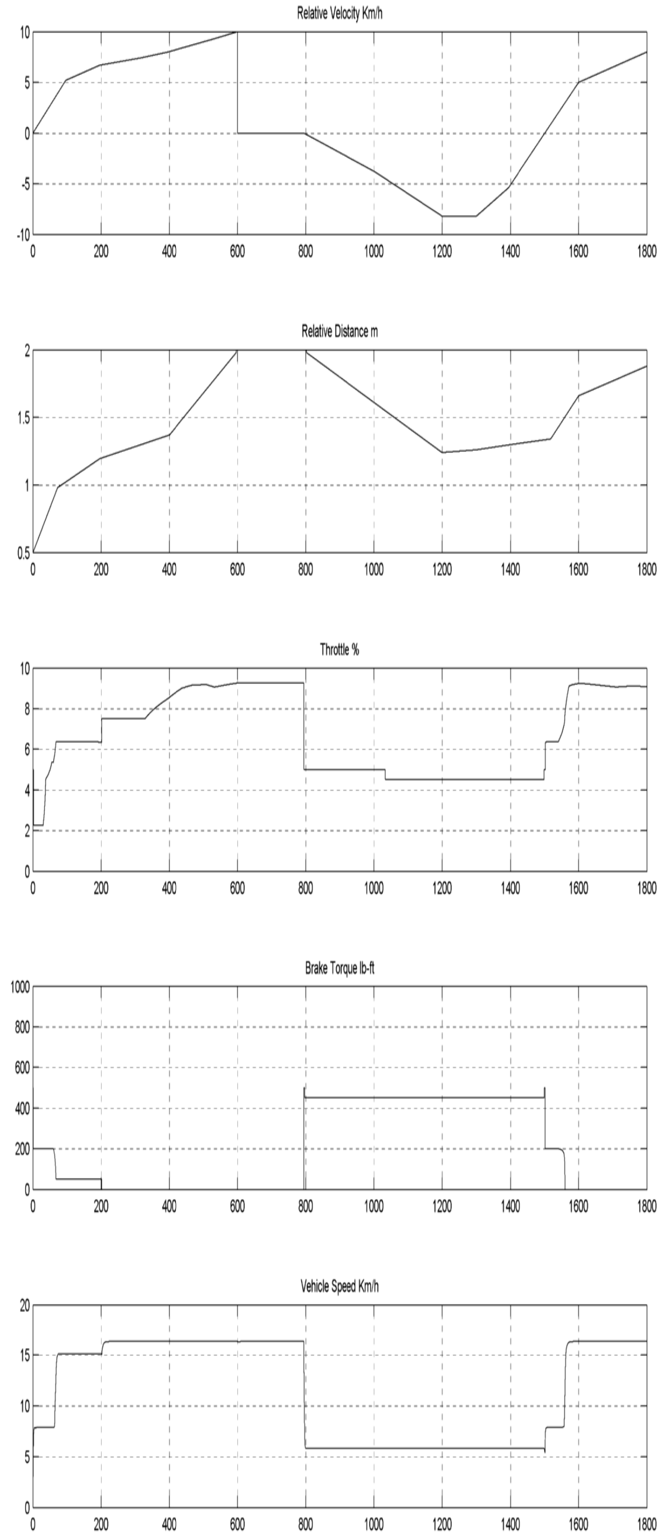


Figure 12. Simulation 4 –Low speed cruise control system case ii

In the last phase of the simulation, the throttle is lowered and the brake torque is increased, since relative velocity becomes negative and the vehicles starts approaching each other. The vehicle decelerates to about 1km/h at the end of the simulation. The fuzzy controller has effectively managed to control the vehicle appropriately throughout the simulation.

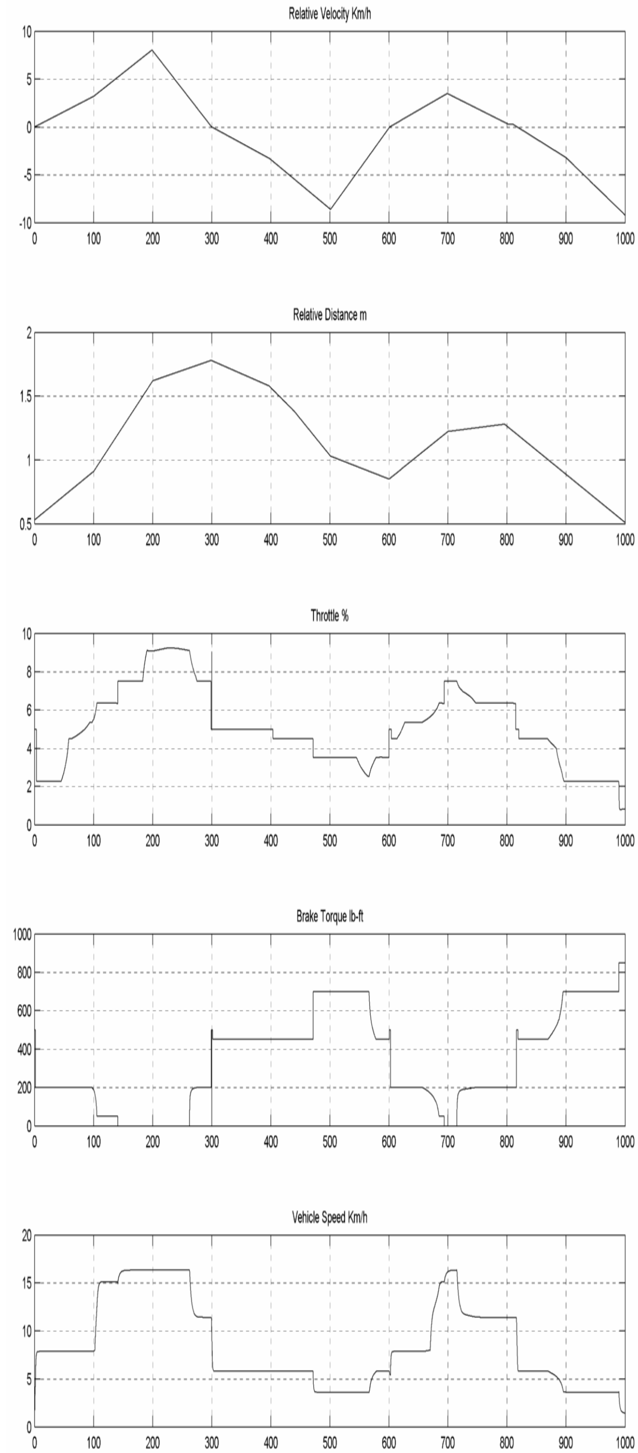


Figure 13. Simulation 5 –Low speed cruise control system case iii

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

D. Simulation 6 –Low speed cruise control system case iv

The vehicle initally accelerates to 7.86 km/h as relative velocity and distance increases and moves at constant speed for another 70s. Then vehicle speed further increases to 16 km/h at 100s. Throttle and brake response during this period can be seen in figure 14. The throttle and brake remain the same till the vehicle remains at this speed

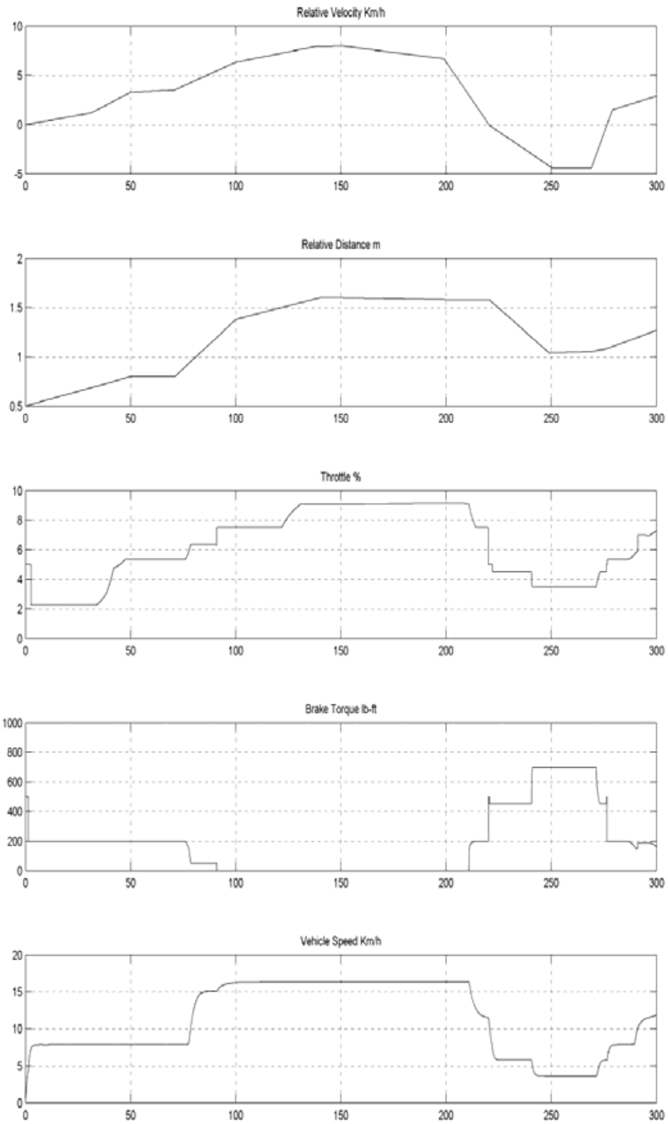


Figure 14. Simulation 6 – Low speed cruise control system case iv

At 225s, relative velocity crosses the zero mark and the vehicle speed begins to decrease. The throttle is lowered further and brake increases in order to lower the vehicle speed further. At 275s, the relative velocity remains constant and therefore, the brake and throttle remain constant. Hence, it can be concluded that the fuzzy controller has responded appropriately when relative velocity is constant.

Finally, the vehicle speed begins to increase as the relative velocity crosses the zero mark. Appropriate changes in the throttle and brake torque can be observed from the figure 14.

E. Simulation7 – Switching of cruise control system to manual mode

This simulation is performed in order to demonstrate that the cruise control system can be switched to manual mode whenever the driver wishes to take over the control of the vehicle.

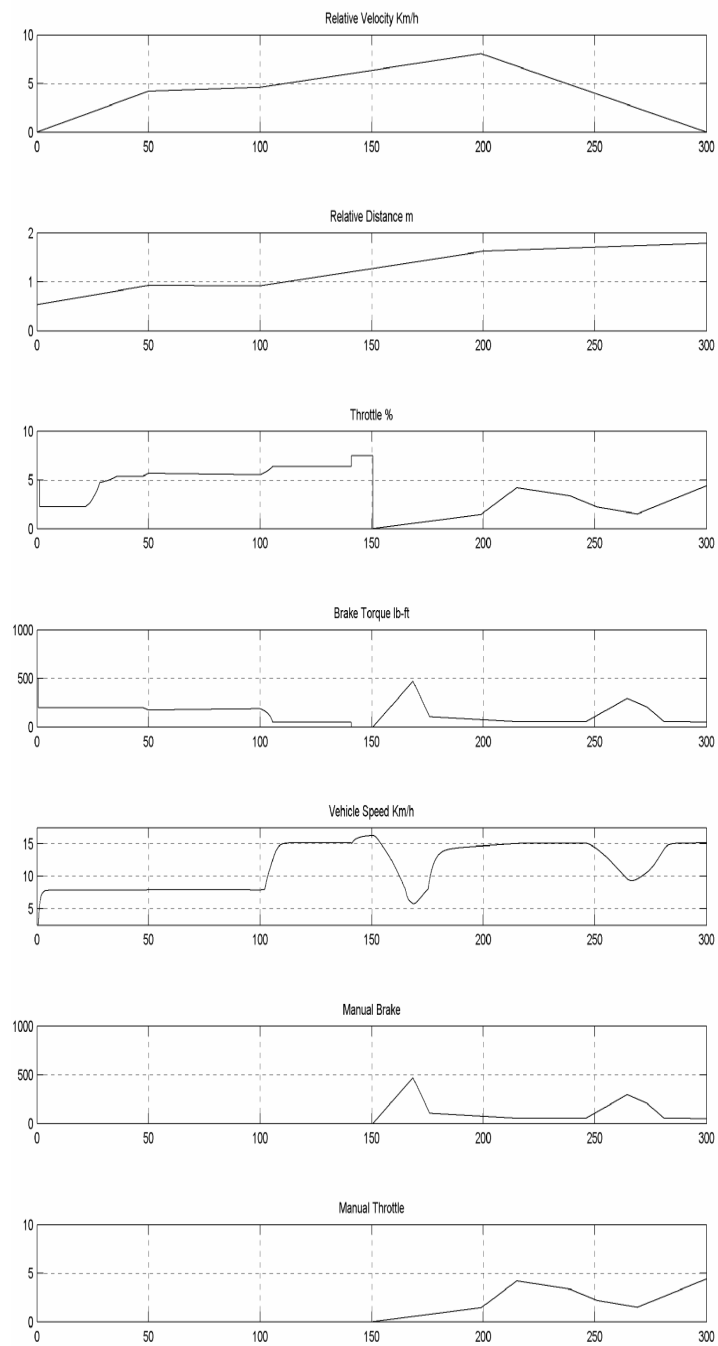


Figure 15. Simulation 7 – Switching of cruise control system to manual mode

The vehicle intially responds to the changes in the relative velocity and distance as it is being controlled. The driver applies the brakes at 150s, as the manual brake graph shown in figure 15. The same brake profile can be seen in brake torque graph. For the next 150s, we can observe from

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

the brake torque graph, that the brake torque profile remains the same as the manual brake profile. This shows that the vehicle is being controlled by the driver. A similar observation can also be made with respect to throttle graph.

This switching of control of the vehicle is performed by adding a switch to the output of the fuzzy logic brake controller and manual brake. The switch has a control input and two data input port. The control input allows the first input, if it is above zero. The brake input has been given both as the control input and the first input. The fuzzy logic brake controller output is given as the second input to the switch. The moment the driver presses the brake pedal, and the brake torque increases above zero, the fuzzy logic controller connection to the vehicle is cut off by the switch and it allows the manual brake input to the vehicle and hence the control of the vehicle is switched to the manual mode. A similar method is applied for the throttle input.

VII. CONCLUSION AND FUTURE WORK

In this project a Fuzzy logic controlled low speed cruise control system is designed successfully and extensive simulation is carried out to test the results..The basis of the design of this control system was to control the vehicle in a congested traffic situation.The inputs to vehicle model are the throttle and brake. Therefore, two separate fuzzy logic controllers were modeled to control these vehicle actuators. To control the throttle, a set of 33 rules were constructed and to control the brake, a set 35 rules were constructed. As both these control variables are fuzzy variables, Mamdani Inference method is chosen to give the output. Two different fuzzy controllers for brake and throttle are used to have better control over the vehicle, by allowing them to operate independently.

In order to maintain a safe distance between the controlled vehicle and the preceding vehicle, two variables are required; the relative velocity and distance between the vehicles. The sensor profile of these variables were assumed from a research report [8]. Using the basis of this profile, different input relative velocity and distance profiles were given to fuzzy logic cruise control system for testing.

The simulation results shows that fuzzy logic cruise control system operates well to control the vehicle effectively based on different input profiles provided to the system.The fuzzy cruise control system completely controls the vehicle without the intervention of the driver. The maximum speed that the vehicle reaches when the cruise control sytem is operating is 16 km/h. This is the typical speed range of vehicles moving in a congested traffic situation.This leads to the conclusion that the fuzzy logic system is able to control the vehicle very well in a low speed heavy traffic environment. Moreover, a switching function has been attached to the system which allows the driver to take over the operation simply by pressing the brake or the throttle without any difficulties.

The drivetrain model used in this system is a very basic one and the vehicle can be controlled in the longitudinal direction only. This low speed cruise control system can be enhanced to a real time system by extending the model of the automobile with feed back sensors appropriately on MATLAB. This will enable the movement of the vehicle in any direction required when simulating the vehicle.

A complete VRML (virtual reality machine language) model for an urban traffic environment shall be designed. This will allow testing of the vehicle in a congested traffic enviroment. The 3D traffic environment will require several external variables such as detection of traffic lights, pedestrians, intersections along with the movement of the vehicle in the congested traffic environment.

The fuzzy logic system must be able to incorporate the environmental conditions when the vehicle is moving in the above 3D model. The VRML realm builder along with the virtual reality toolbox allows the connection of the fuzzy cruise control system to the 3D system. This will make the testing of the system more realistic.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Dr. M. Ramachandran – Director, Bits Pilani, Dubai for granting me an opportunity to perform this project in the college environment and use its resources. I would also like to thank my thesis guide Dr.Mary Lourde R for her constant guidance and support, which has helped me in making this project.

REFERENCES

[1] Toyota Yaris Sedan Specification , www.thecarconnection.com 2007 Toyota Yaris 4dr Sedan Auto S (Natl)
http://www.thecarconnection.com/specifications/toyota_yaris_2007_4dr-sdn-auto-s-se_performance-specs

[2] R. Garcia et al., “Frontal and Lateral Control for Unmanned Vehicles in Urban Tracks,”*IEEE Intelligent Vehicle Symp*. (IV2002), vol.2, IEEE Press, 2002, pp. 583–588.

[3] José E. Naranjo, et al., “Using Fuzzy Logic in Automated Vehicle Control” IEEE INTELLIGENT SYSTEMS, January 2007:36-44

[4] M.A. Sotelo et al., “Vehicle Fuzzy Driving Based on DGPS and Vision,” *Proc. 9th Int’l Fuzzy Systems Assoc.*, Springer, 2001, pp.1472–1477.

[5] A. Broggi et al., “The Argo AutonomousVehicle’s Vision and Control Systems,” *Int’l J. Intelligent Control and Systems*, vol. 3, no.4, 2000, pp. 409–44.

[6] Using Simulink and Stateflow in Automotive Applications, www mathworks.com Modeling an Automatic Transmission Controller
http://www.mathworks.com/products/simulink/demos.html?file=/products/demos/shipping/simulink/sldemo_autotrans.html

[7] K. Naab et al., “Stop and Go Cruise Control”, “International Journal of Automative Technology”, Vol 1,No .2, June 2000, pp 61-69

[8] K. Hedrick, J. Jang, and A. Potier, "Cooperative Multiple-Sensor Fusion for Automated Vehicle Control" (April 1, 2004). California Partners for Advanced Transit and Highways (PATH)

[9] Fancher et al.,“Intelligent Cruise Control Field Operational Test (Final Report)”,University of Michigan Transportation Research Institute (1998).

[10] Ing. Ondřej Láník, “Fuzzy Logic Vehicle Intelligent Cruise Control Simulation”, Czech Technical University in Prague, Faculty of Mechanical Engineering,Technická 4, CZ – 166 07 Praha 6, Czech Republic.

[11] Micheal Klotz, Rohling H, “24 Ghz for Automotive applications”, Journal of Telecommunications and Technology, April 2001, pp 11-14.

Plant Classification Based on Leaf Recognition

Abdolvahab Ehsanirad
Department of Computer Science
Islamic Azad University, Minoodasht Branch, Iran
Email: vahab61@gmail.com

Abstract — **Current study used the image processing techniques in order to classification of plants based on leaves recognition. Two methods called the Gray-Level Co-occurrence matrix (GLCM) and Principal Component Analysis (PCA) algorithms have been applied to extract the leaves texture features. To classify 13 kinds of plants with 65 new or deformed leaves as test images, the Algorithms are trained by 390 leaves. The findings indicate that the accuracy of PCA method with 98% come out to be more efficiency compare to the GLCM method with 78% accuracy.**

Keywords - Classification, GLCM, PCA, Feature Extraction.

I. INTRODUCTION

Leaf recognition is a pattern recognition task performed specifically on leaves. It can be described as classifying a leaf either "known" or "unknown", after comparing it with stored known leaves. It is also desirable to have a system that has the ability of learning to recognize unknown leaves.

Computational models of leaf recognition must address several difficult problems. This difficulty arises from the fact that leaves must be represented in a way that best utilizes the available leaf information to distinguish a particular leaf from all other leaves.

Compared with other methods, such as cell and molecule biology methods, classification based on leaf image is the first choice for plant classification. Sampling leaves and photogeny them are low-cost and convenient. One can easily transfer the leaf image to a computer and a computer can extract features automatically in image processing techniques. Some systems employ descriptions used by botanists. But it is not easy to extract and transfer those features to a computer automatically. It is difficult job to tell the just one algorithm alone is the best and successful at recognizing any and all variation of the same object. And it is more difficult to tell the same algorithm to be able to differentiate between different objects. Many research has done for the leaf classification with some texture feature extraction methods [3,9,10,7].

II. LEAF CLASSIFICATION PROCESS METHOD

The conventional method of leaf classification involves two main steps. The first step is obtaining a priori knowledge of each class to be recognized. Normally this knowledge encompasses some sets of texture feature of one or all of the classes. Once the knowledge is available and texture feature of the observed image are extracted, then classification techniques, for example nearest neighbors and decision trees, can be used to make the decision [5], that is the second step. Such a procedure is illustrated in Figure 1, the tasks that texture classification has been applied to include the classification of plant leaves images [2].

Currently there are a huge number of texture feature extraction methods available and most of the methods are associated with tunable parameters. It is difficult to find the most suitable feature extraction methods and their optimal parameters for a particular task. In addition, performance of classification methods also depends upon the problems, which makes selecting an optimal "feature extraction + classification" combination a difficult assignment.

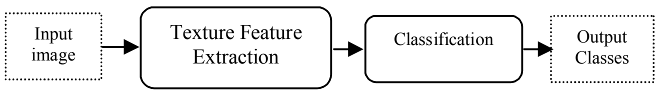


Figure 1. Conventional Plant Classification Process

III. FEATURE EXTRACTION

Different features are chosen to describe different properties of the leaves. Some leaves are with very distinctive shape, some have very distinctive texture patterns, and some are characterized by a combination of these properties.

IV. TEXTURE ANALYSIS

Texture analysis mainly aims to computationally represent an intuitive perception of texture and to facilitate automatic processing of the texture information for artificial vision systems. The process of texture analysis usually produces kind of numeric descriptions of the texture, called texture features. The process of computing the texture feature is known as feature extraction.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

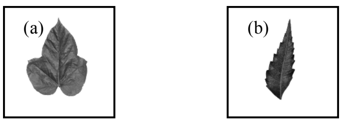
There are an enormous number of texture analysis methods under this category although none predominates. Methods that we used for classification will describe here.

V. GRAY-LEVEL CO-OCUURRENCE MATRIX

This method was first proposed by Haralick in 1973 and still is one of the most popular means of texture analysis [8]. The key concept of this method is generating features based on gray level co-occurrence matrices (GLCM). The matrices are designed to measure the spatial relationships between pixels. The method is based on the belief that texture information is contained in such relationships.
Co-occurrence features are obtained from a gray-level co-occurrence matrix. We used 22 features that extracted from GLCM matrix in our paper [8,4,1].

VI. TEXTURAL FEATURES EXTRACTED FROM GRAY-LEVEL CO-OCUURENCE MATRICES

Our initial assumption in characterizing image texture is that all the texture information is contained in the gray-level Co-occurrence matrices. Hence all the textural features here are extracted from these gray-level Co-occurrence matrices. The equations which define a set of 22 measures of textural features are given in this paper. Some GLCM Extracted textural features are illustrated in Table 1 for two different leaf images.



some texture Features extracted from Leaf image (a)					
Angle	Autocorrelation	Entropy	Contrast	Correlation	Homogeneity
0°	45.5748	1.4311	0.3184	0.9638	0.6144
45°	45.2799	1.4928	0.4458	0.9496	0.6060
90°	45.6190	1.3886	0.2301	0.9738	0.6166
135°	45.2932	1.4716	0.4192	0.9526	0.6074

some texture Features extracted from Leaf image (b)					
Angle	Autocorrelation	Entropy	Contrast	Correlation	Homogeneity
0°	54.8540	0.8972	0.4361	0.9401	0.8214
45°	54.7371	0.9132	0.4438	0.9396	0.8199
90°	54.9797	0.8405	0.1845	0.9747	0.8267
135°	54.6610	0.9310	0.5961	0.9189	0.8172

Table 1. GLCM Extracted textural features for two different leaf images.

VII. LEAF CLASSIFICATION USING PCA

In this study, we have followed the method which was proposed by M. Turk and A. Pentland [6] inorder to develop a leaves classification system based on the eigenspace approach. They argued that, if a multitude of leaf images can be reconstructed by weighted sum of a small collection of characteristic features or eigenpictures, perhaps an efficient way to learn and recognize leaves would be to build up the characteristic features by experience over time and recognize particular leaf by comparing the feature weights needed to approximately reconstruct them with the weights associated with known leaves. Therefore, each leaf is characterized by a small set of feature or eigenpicture weights needed to describe and reconstruct them. This is an extremely compact representation when compared with the images themselves.

VIII. EXPERIMENTAL RESULTS AND DISCUSSINONS

The experiment is designed to illustrate the performance of two feature extraction methods, GLCM and PCA algorithms for plant leaves classification purpose.
The GLCM is a tabulation of how often different combinations of pixel brightness values (grey levels) occur in an image. The classification steps are illustrated in Figure 2.
In the first experiment after changing the color image to gray-level image with using of the GLCM texture feature extraction we extracted the 22 features [8,4,1] of each leaf images.

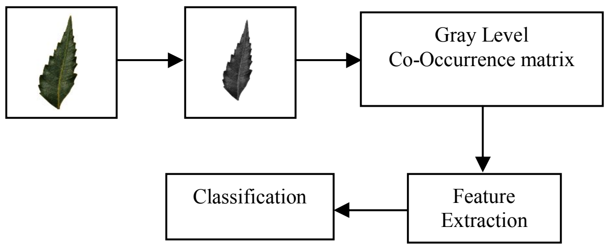


Figure 2. Classification Steps in GLCM method.

Features	Sample leaf from leaves Classes					
	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
Autocorrelation	55.27373	49.56994	60.29225	54.31234	45.25222	50.65949
Contrast	0.332278	0.368038	0.322468	0.306646	0.310127	0.333228
Correlation	0.955264	0.961478	0.87063	0.956296	0.961898	0.959497
Dissimilarity	0.126899	0.174367	0.123101	0.138608	0.172468	0.151266
Energy	0.699192	0.528028	0.828267	0.646948	0.387419	0.549963
Entropy	0.817156	1.286197	0.574364	1.034476	1.52748	1.168368

Table 2. Some features extracted from some chosen leaf image of each leaves classes in (d=1) and degree 0°.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

We have tried the GLCM method with Distance 1 (d=1) and degree 0°, Distance 1 (d=1) and degree 45°, Distance 1 (d=1) and degree 90° and Distance 1 (d=1) and degree 135°. The performance accuracy of each one is shown in Table 3.

In our experience, GLCM method in leaf recognition for the degrees 0° and 90° gave the same accuracy and same result. Here the poor result is in the 45° degree. Because any changes in the neighboring distance or the neighboring degree it will change the value of extracted texture feature.

degree	Average recognition rate (%)
0°	78.46
45°	49.23
90°	78.46
135°	70.76
PCA	98.46

Table 3. The performance of GLCM method in different degrees with neighborhood distance 1 and performance of PCA method.

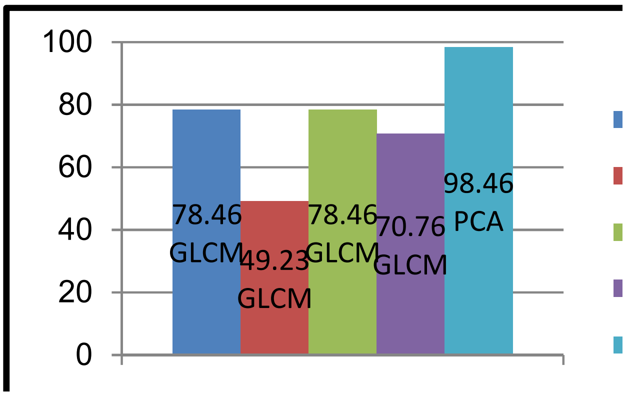


Figure 3. PCA and GLCM accuracy chart in different degrees

The GLCM method is very sensitive for the any changes in the images such rotation, scale and *etc*. In (Tables 2) you can see the some in extracted features in neighborhood degree 0. The computation time for GLCM method is less and recognition of this method is very fast.
PCA method mostly using for the face recognition purpose but we tried as leaf recognition. In PCA also image should be change to gray level that can reduce the image dimension. In our experience the PCA method gave the efficient performance and very good result. It was the just one wrong recognition out of 65 test image in our test. But the test speed is not much good and computation time is high for recognizing one test image. Compare with GLCM it's very slow but the performance of PCA method is efficient (Figure 3).

IX. Database

The database used in our experiment is collected by our self. We pluck the leaf from the plant in the fields near our campus and around University of Mysore, which consists of intact and fresh leaf images in different rotation for 13 plant species class and constructed by our self. We taken 390 images as training set and each plant class contains the 30 leaf images in different degree of rotation and different leaf images. The test set contains the 65 of deformed and new leaf images and for each class has 5 leaf images for test. The sample dataset of leaf images and related classes are illustrated in Figure 4.

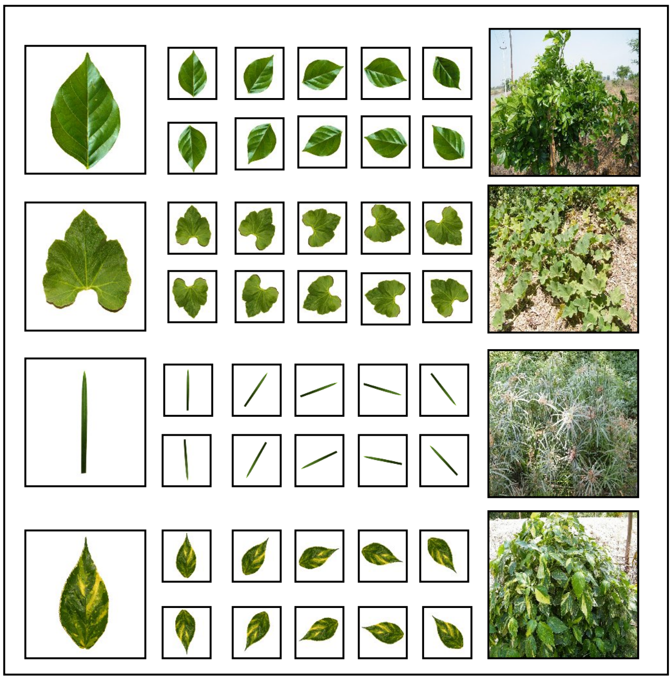


Figure 4. The sample dataset of leaf images and related classes

X. Conclusion

In this study, the classification based on the recognizing the leaves images with extracted texture features was proposed and performed. The texture features have been extracted with using the GLCM and the PCA algorithms, on the 390 image in dataset and with 65 deformed or new leaf images for test. In addition, different degrees for the GLCM method were used and it was found out to be more efficient in the degree 0° by 78.46 % accuracy. Therefore, it was specified that the GLCM is very sensitive in any changes for images such as deforming or giving the new leaf image as a test. In addition, the PCA method comes out to be more efficient compare to the GLCM method by 98.46 % accuracy. Considering the time of

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

recognizing an image as one of the main criteria for classification, the study found out that the GLCM method by taking just 5" second for any test is far better than the PCA method which takes more than one minute (1':6"). Furthermore, the calculation time in the PCA method is time consuming for example making the Eigenvector from considered leaves dataset almost took 2 hours and it was just for 390 images. However, making the dataset images vectors it is for one time and it will not be the big problem in recognizing process.

Moreover, in the future works researchers can either use more images or other methods in order to compare the results of current study with their results.

REFERENCES

[1] D. A. Clausi, "An analysis of co-occurrence texture statistics as a function of grey level quantization", *Can. J. Remote Sensing*, Vol. 28, No. 1, pp. 45–62, 2002.

[2] F. Dell' Acqua and P. Gamba. Texture-based characterization of urban environments on satellite sar images. *IEEE Transaction on Geoscience and Remote Sensing*, 41(1):153-159, January 2003.

[3] j. Graham, "Application of the Fourier-Mellin transform to translation-, rotation and scale-invariant plant leaf identification" McGill University, Montreal, July 2000.

[4] L. K. Soh and C. Tsatsulis. Texture Analysis of SAR Sea Ice Imagery Using Gray Level Co-Occurrence Matrices. In *IEEE Transaction on Geoscience and Remote Sensing 1999,* volume 37, No 2, March 1999.

[5] M. Tuceryan and A.K. Jain. Texture analysis. In C. H Chen, L. F. Pau, and P. S. P. Wang, editors, *Handbook of Pattern Recognition and Computer Vision*, chapter 2, pages 235-276. World Scientific, Singapore, 1993.

[6] M. Turk, A. Pentland, "*Eigenfaces for Recognition*", Journal of Cognitive Neuroscience, Vol. 3, No. 1, 1991, pp. 71-86.

[7] P. Tzionas, S. E. Papadakis, D. Manolakis, "Plant leaves classification based on orphological features and a fuzzy surface selection technique ",in *Fifth International Conference on Technology and Automation,* Thessaloniki, Greece, 2005, pp. 365-370.

[8] R. M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3(6):610-621, Noveber 1973.

[9] S. G. Wu, F. S. Bao, E. Y. Xu, Y. Wang, Y. Chang and Q. Xiang, "A Leaf ecognition Algorithm for Plant Classification Using Probabilistic Neural Network" arxiv, 0707, 4289v1, [cs.AI], 29 Jul 2007.

[10] Z. Wang, Z. Chi, D. Feng and Q. Wang, "Leaf Image Retrieval with Shape Features", R. Laurini (Ed.): VISUAL 2000, LNCS 1929, pp. 477–487, 2000.

Abdolvahab Ehsanirad received the B.E. in Computer Software Engineering degree in 2006 from Islamic Azad University of Sari, Iran, and M.Tech in Computer Science and Technology degree in 2010 from University of Mysore, India. His interest research areas are image processing and pattern recognition. At present he is doing some research in image processing at Islamic Azad University, Minoodasht branch, Iran.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

RELIABLE ROUTING WITH OPTIMIZED POWER ROUTING FOR WIRELESS ADHOC NETWORK

T.K.Shaik Shavali , Dr T. Bhaskara Reddy and Sk fairooz

Professor , Department of Computer Science, Lords institute of Engineering & Tech,

Hyderabad-08, A.P. , INDIA

E-mail:- ssvali786@yahoo.com

Department of Computer Science & Technology, S.K. University, Anantapur-03, A.P.,INDIA

E-mail:-bhaskarreddy_sku@yahoo.co.in

Associate Prof, Department of ECE, AHCET, Hyderabad-08, A.P. , INDIA

E-mail:-fairoozsk@gmail.com

Abstract

In this paper work, a routing protocol called RMP (route management protocol) is implemented to cope with misbehavior operation in AdHoc network. It enables nodes to detect misbehavior by first-hand observation and use the second-hand information provided by other nodes. This RMP protocol can run on any routing protocol to cope with misbehavior. In this paper work, we have tested for DSR routing protocol.(ie DSR with RMP). The efficiency of communication routes is tested over the node power consumption and developed a mechanism to optimize the power consumption in routing scheme.

Keyword: route management protocol, adhoc network, power optimization, network efficiency

I. INTRODUCTION

Wireless networking grows rapidly because of the human desires for mobility and for freedom from limitation, i.e., from physical connections to communication networks. A particular kind of wireless network called mobile ad hoc networks is presently under development. A mobile ad hoc network is a self-organizing and rapidly deployable network in which neither a wired backbone nor a centralized control exists. The network nodes communicate with one another over scarce wireless channels in a multi-hop fashion. The ad hoc network is adaptable to the highly dynamic topology resulted from the mobility of network nodes and the changing propagation conditions. These networks are used in emergency disaster rescue operation, tactical military communication and law enforcement.. Mobile ad hoc networks are also a good alternative in rural areas or third world countries where basic communication infrastructure is not well established

The lack of infrastructure and organizational environment of mobile ad hoc networks offer special opportunities to attackers. Misbehavior means deviation from normal routing and forwarding behavior. Without appropriate countermeasures, the effects of misbehavior dramatically decrease network performance. Depending on the proportion of misbehaving nodes and their specific strategies, network throughput can be severely degraded, packet loss, nodes can be denied service, and the network can be partitioned. These detrimental effects of misbehavior can endanger the functioning of the entire network. Minimizing energy consumption is the important challenge in mobile networking. Wireless network interface is often a device's single largest power consumer. Since the network interface may often be idle, turning the radio off when not in use could save this power. In practice, however, this approach is not straightforward. A node must arrange to turn its radio on not just to send packets, but also to receive packets addressed to it and to participate

in any higher-level routing and control protocols. The requirement of cooperation between power saving and routing protocols is particularly acute in the case of multi-hop ad hoc wireless networks, where nodes must forward packets for each other.

II. POWER OPTIMIZED ROUTING SCHEME

The Topology Management in Ad hoc Wireless Networks is deciding for every node:

1.which node to turn on.

2.when they turn on.

3.At what transmit power.

In power on-off scheduling topology management schemes, few nodes, rich in power, are selected as cluster heads and gateways. These cluster head nodes are selected distributive in such a way that each node in the ad hoc wireless network is either cluster head or connected (i.e., in transmission range) to the cluster head and the gateway nodes are selected such that they forward packets between cluster heads. Cluster heads and gateways form the virtual backbone for routing in ad hoc networks. Some proposed power on-off scheduling topology management schemes are Span (3) and TMPO (Topology Management by Priority Ordering) (4). In span, some special coordinator node are selected distributive in such a way that two of the coordinators neighbors can not reach each other either directly or via one or two coordinators. This selection rule ensures the connectivity in ad hoc network. Span runs over 802.11 ad hoc power saving mode, which has high broadcast overhead. While TMPO assigns willingness value to each node, based on the energy level and speed of the node. A node with high willingness value is selected as cluster head with high probability.

Few power scheduling topology management schemes are CBTM (Cone based Distributed Topology Management) and K-Neigh Protocol for symmetric topology control (6). In CBTM, each node tries to find the minimum transmitting power p such that transmitting with p ensures that in every cone of degree around each node, there is at least one neighbor node. Whereas in the K-Neigh Protocol, each node adjusts its transmission power, such that it has k or slightly less than k one-hop neighbors, So that network connectivity is maintained under the conditions of mobility. Most of the algorithms proposed for Topology Management follow either first two steps or third step, i.e., switching between active (transmit, receive or idle) and sleep mode or the second step, i.e. adjusting the transmission power. We call them as power on-off scheduling and transmit power-scheduling algorithms respectively.

Objective of this work is to design a topology management scheme for ad hoc wireless networks. A good power-saving topology management scheme for wireless ad hoc networks should have the following characteristics:

It should allow as many nodes as possible to turn their radio receivers off most of the time because even an idle radio in receive mode can consume almost as much energy as an active transmitter.

The algorithm for picking this backbone should be distributed, requiring each node to make a local Division.

III. PROPOSED TOPOLOGY MANAGEMENT SCHEME

In our topology management scheme, power (Mobile Agent with Routing Intelligence) nodes are selected in such a way that power nodes have the maximum power level among their on hop neighbors and all non-power nodes are within the transmission range of power nodes. These power nodes have the routing intelligence i.e. they make all decisions related to routing. The gateway nodes having sufficient power level are selected so that they can forward packets between power nodes. A gateway node does not have routing intelligence. These power and gateway nodes stay continuously awake to route the packets of other member nodes. The member nodes wake up a number of times in a beacon period T, and if they do not have to transmit or receive data, they go to sleep mode again. The wake up time for each node is calculated from a pseudo-random number, such that power node and neighbor nodes know the wake up of that node time.

Thus the member node can remain in power saving sleep mode most of the time, if it is not actively sending or receiving packets. The packets are routed over the virtual backbone consisting of power nodes and gateways. The routes are found with the help of mobile agents.

The topology management scheme runs above the MAC layer and interacts with the routing protocol. If a node has been asleep for a while, packets destined for it are not lost but are buffered at a neighboring power node. When the node awakens, it can retrieve these packets from the buffering power node. This topology management schemes makes the routing simple, as only those entries in a node’s routing table that correspond to

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

currently active power nodes can be used as valid next-hops (unless the next hop is the destination itself).

Definition 1 Power nodes are the nodes such that all non-power nodes are connected to (i.e., in transmission range of) power nodes and route packet for all other nodes with the help of mobile agents.

Definition 2 Sleep Cycle period is the time period during which member nodes remain in the power efficient sleep mode and wake up once for fixed time duration T.

We assume that each node periodically broadcasts HAI messages that contains:

Node's id,

Its status (i.e., whether the node is a power node, gateway, member, undecided),

Its current power level,

Its current power node,

A wakeup counter wi,

Information about each neighbor i.e.

Neighbor's id,

Its status,

Its power node.

Based on the HAI messages received from neighbors, each node constructs a list of its neighbors, their power nodes, power level, wakeup counter and information about their neighbors.

A node switches state from time to time between being a power node and being a member. A node becomes a gateway, if its power node chooses it as a gateway to route the packets between power nodes. It switches its state to undecided, if it looses contact with its power node due to mobility. A node includes its current state in its HAI messages. The following sections describe that it should withdraw from being a power node, and how a power node selects its gateways.

a) POWER MANAGEMENT

Power nodes along with gateways form the virtual backbone, which is used for routing this demands for additional power for transmission, reception and processing of routing packets. Thus these power nodes should be selected in such a way that they have enough power level

The nodes in a the network periodically check among its one hop neighbors for maximum power and declare themselves a node as power if it has maximum power. Power nodes select the neighboring nodes as its members and maintain the list about its members. If more than one neighbors of an undecided node become power then undecided node selects its power node from which it has received the HAI packet earlier. If an undecided node has more power than power node then it declares it self as power in the next HAI packet.

b) POWER NODE WITHDRAWAL

Power node will drain its energy more rapidly, as compared to member nodes. Before the power node loses its major part of its power, responsibility of power node should be transferred to other node with sufficient power level. Also power nodes should not be changed frequently which will increase the overhead.

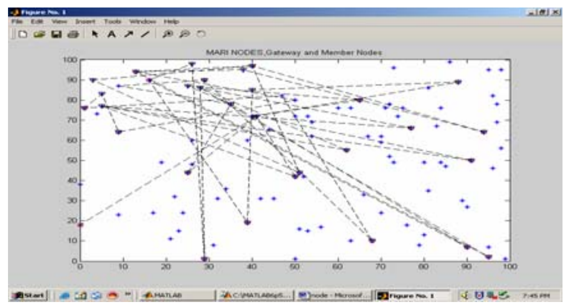


Fig 1 nodes distributed in an network

c) GATE WAY SELECTION

As the maximum number of hops between any two close power nodes is two, gateways are required forward packets between power nodes. Also as gateways need to receive and transmit routing packets to and from power nodes, they should have sufficient amount of power.

Power nodes periodically send broadcast request packet STAY-AWAKE to its members for synchronization among members. Then it selects a node as gateway, which has maximum power and maximum power nodes as its neighbors. If any power nodes with in two hops have already declared their gateways, then there is no need to select gate way again. Power level of the gateway is periodically checked by power and if it has less power than threshold, then power starts new gateway selection.

d) SLEEP CYCLE SCHEDULING:

We propose some additional power saving features to 802.11 CSMA/CA to make the MAC layer power efficient by using randomized wake up time for member nodes in ad hoc network. Power nodes and gateways continuously stay awake to forward packets of other nodes. Member nodes wake up a number of times in a beacon period T (see figure) and if they do not have to transmit or receive data, they

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

go to sleep again. There are number of sleep cycle periods (T1, T2), (T2, T3)…(Tn, T) in a beacon period. Member nodes wake up once in a sleep cycle. All nodes stays awake during period (0,T1) called as broadcast window to exchange HAI packets. Each node synchronizes their clock by using time stamp of HAI message from power node. Each member node determines its wake up time from its node id and a wakeup counter wi

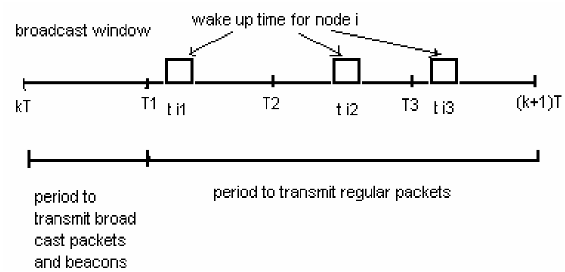


Fig 2 beacon period

e) LOAD DISTRIBUTION

One part of ad hoc network may be congested and other part of network may have free resources. This will increase the packet delivery latency. Throughput and packet delivery ratio also will be badly affected. To distribute the load evenly in the network, we have devised a congestion metric, which is used for route selection as described above. This congestion metric is based on the amount of time power node sees free channel for the past T seconds.

The developed power optimization scheme is then incorporated with a routing scheme for the reduction of non-cooperative nodes so as to minimize the power consumption happing at each node. The approach of discovering and reduction of misbehaving nodes are as outlined below.

IV. ROUTING PROTOCOL

Dynamic Source Routing is a protocol developed for routing in mobile ad-hoc networks. Nodes send out a ROUTE REQUEST message, all nodes that receive this message forward it to their neighbors and put themselves into the source route unless they have received the same request before. If a receiving node is the destination, or has a route to the destination, it does not forward the request, but sends a REPLY message containing the full source route. It may send that reply along the source router in reverse order or issue a ROUTE REQUEST including the route to get back to the source, if the former is not possible due to asymmetric links ROUTE REPLY messages can be triggered by ROUTE REQUEST messages or gratuitous. After receiving one or several routes, the source picks the best (by default the shortest), stores it, and sends messages along that path. In general, the better the route metrics (number of hops, delay, bandwidth or other criteria) and the sooner the REPLY arrived at the source (indication of a short path - the nodes are required to wait a time corresponding to the length of the route they can advertise before sending it in order to avoid a storm of replies), the higher preference is given to the route and the longer it will stay in the cache. In case of a link failure, the node that cannot forward the packet to the next node sends an error message toward the source. Routes that contain a failed link, can be 'salvaged' by taking an alternate partial route that does not contain the bad link.

a) PASSIVE ACKNOWLEDGMENT (DSR)

Instead of waiting for an explicit acknowledgment for each packet by the next-hop node on the route, a node assumes the correct reception of the packet when it overhears the next-hop node forwarding the packet this is called passive acknowledgment In this, the simple passive acknowledgment is used not only for an indication of correct reception at the next hop, but also to detect if nodes fail to forward packets.

b) MISBEHAVIOR CLASSIFICATION BY DSR

DSR classify attacks on them as dropping, modification, fabrication, or timing attacks. The previous hop can detect dropping by use of passive acknowledgment and this is detected as misbehavior and takes an alternate partial path to reach to the destination.

V. RMP PROTOCOL

a) RMP Protocol Components

We present here the RMP components we designed for coping with routing and forwarding misbehavior in mobile ad-hoc networks running DSR Fig 3 shows the protocol components.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

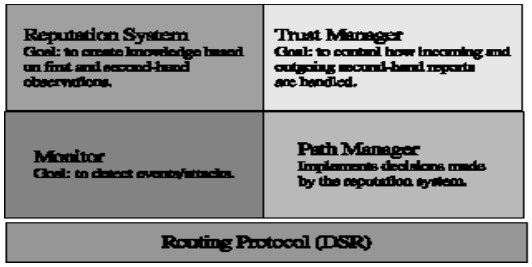


Fig 3: RMP protocol components

b) MONITOR

The goal of the monitor is to gather first-hand information about the behavior of nodes in the network. This is achieved by observing and classifying node behavior as normal or misbehaving. The monitor can detect misbehavior that can be distinguished from normal behavior by observation. We call the information gained by direct experience by node i about node j first hand information (Fi,j) and use it as an input to the reputation system component of RMP.

c) REPUTATION SYSTEM

Reputation systems are used for example in some online auctioning systems. The main idea behind the use of reputation systems is two fold first; it is used to serve as an incentive for good behavior to avoid the negative consequences that a bad reputation can entail. Second, it provides a basis for the choice of prospective transaction partners. The most relevant properties of a reputation system are the representation of reputation, how the reputation is built and updated, and for the latter, how the ratings of others, i.e. second-hand information, are considered and integrated. The reputation of a given node is the collection of ratings maintained by others about this node. In our approach the reputation system is fully distributed, and a node i maintains ratings about every other node j that is cares about. The reputation rating represents the opinion formed by node i about node j's behavior as an actor in the base system, i.e. whether node j correctly participates in the routing protocol and forwarding. We represent the reputation ratings that node i has about node j as data structure Ri,j

The use of second-hand information enables nodes to find out about misbehaving nodes before making a bad experience. Also, in mobile ad-hoc networks, nodes might not meet every node that they need for multi-hop forwarding, but with second-hand information they can make informed decisions about which node to use for their paths.

d) TRUST MANAGER

The task of the trust manager is to decide when to trust second-hand information and to administer the trust given to other nodes. The goal is to minimize the risk of spurious ratings while still making use of second-hand information received from others. The trust rating represents node i's opinion about how honest node j is as an actor in the reputation system (i.e. whether the reported first hand information summaries published by node j are likely to be true). We represent the trust ratings as data structure Ti,j

e) PATH MANAGER

Once a node i classifies another node j as misbehaving, i isolates j from communications by not using for routing and forwarding and by not allowing j to use i .This isolation has three purposes. The first is to reduce the effect of misbehavior by depriving the misbehaving node of the opportunity to participate in the network. The second purpose is to serve as an incentive to behave well in order not to be denied service. Finally, the third purpose is to obtain better service by not using misbehaving nodes on the path. The path manager performs the following functions: Path re-ranking according to security metric (e.g. reputation of the nodes in the path), deletion of paths containing misbehaving nodes, action on receiving a request for a route from a misbehaving node (e.g. ignore, do not send any reply), and action on receiving request for a route containing a misbehaving node in the source route (e.g. ignore, alert the source). The path manager thus controls the topology as seen by an individual node. Misbehaving nodes are not used for routing and forwarding and the path manager refuses to be used by them.

V. RESULTS

1) Fig 4 shows a dynamic ad hoc network with 20 nodes and 8 misbehaving nodes distributed randomly.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

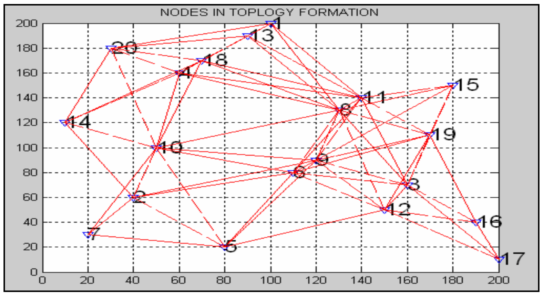


Fig 4: Ad hoc network with 20 nodes and 8 misbehaving nodes.

2) The Fig 5 shows the performance of DSR with RMP protocol for the above randomly distributed network we have chosen 7 as source node and 15 as destination node. Misbehaving nodes are indicated by round circles .The black dotted line shows the optimum path that has been selected to reach the destination. The Fig clearly shows that RMP protocol is able to cope with misbehavior in mobile ad hoc networks thus making network function for normal nodes when other nodes don't route and forward correctly. The protocol is integrated with modified Bayesian approach to decide whether node is misbehaving or not

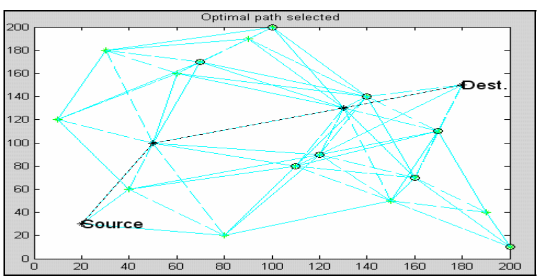
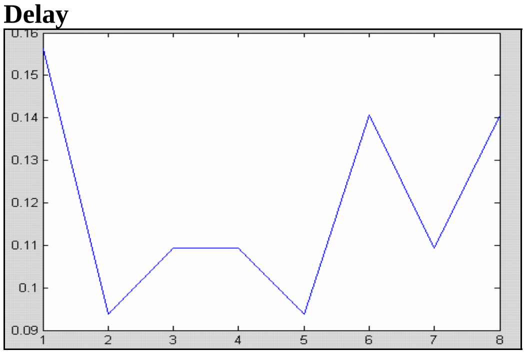


Fig 5: DSR with RMP performance for the randomly distributed network.

3) Fig 6 shows DSR delay plot to reach the destination .As the no of misbehaving nodes goes on increasing the delay to reach the destination increases.

Delay

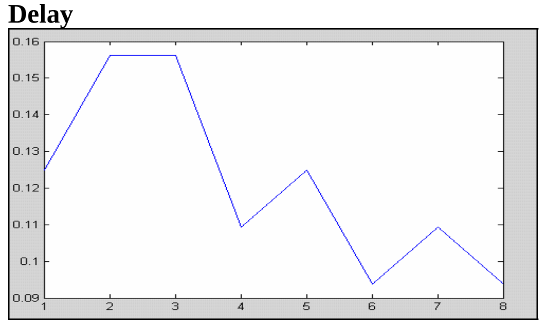


No of misbehaving nodes

Fig 6.DSR delay plot with no of misbehaving nodes

4) Fig 7 shows that our approach of using second hand information not only speeds up the detection of misbehaving nodes but also reduces delay to reach the destination i.e. as the no of misbehaving nodes goes on increasing delay reduces than compared to original DSR.

Delay

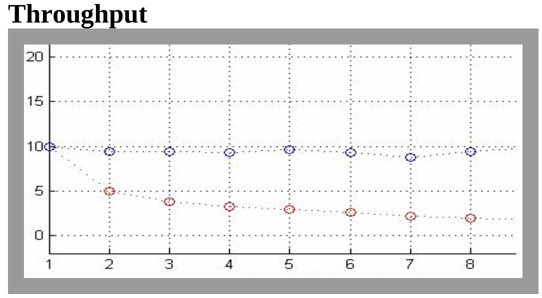


No of misbehaving nodes

Fig 7 DSR with RMP delay plot with no of misbehaving nodes

5) Fig 8 shows the throughput comparison between original DSR and DSR with RMP. The Fig shows that as the number of misbehaving nodes increases throughput decreases in case of original DSR where as DSR with RMP will maintain the constant throughput. RMP can keep the network throughput constant up to 80% misbehaving nodes.

Throughput



No of misbehaving nodes

Fig 8:Throughput comparison between DSR and DSR with RMP

We have used the energy consumption model of [8], which is obtained from measurements on the Cabletron Roam about 802.11 DS High Rate network interface card (NIC) operating at 2 Mbps. Power consumption in various modes such as Tx (transmit), Rx (receive), Idle and sleeping.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Table 1: Power consumption in various modes

Tx	Rx	Idle	Sleeping
1400mW	1000mW	830mW	130mW

The measure the effectiveness of the Topology Management scheme, we simulated, with on demand routing, several static and mobile topologies. Simulation results show that the scheme performs well by low packet delivery latency and high percentage of packet delivery. It outperforms flat topology network (network without using topology management) in average power consumption per node and network lifetime.

a) Fraction of nodes in forwarding backbone

Figure 9 shows the fraction of nodes that are part of virtual forwarding backbone (i.e. RIMA and gateway nodes) as node density increases. It can be observed that as node density increases, fraction of forwarding nodes goes on decreasing. Thus more number of nodes are member nodes, which are in power efficient sleep state most of the time.

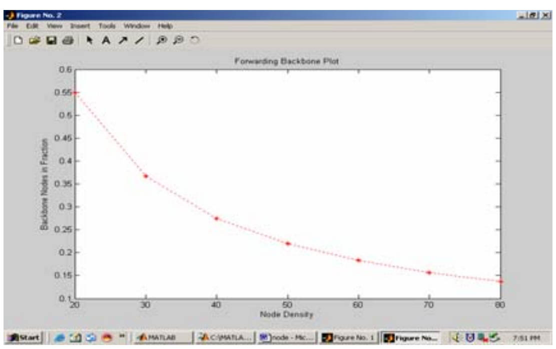


Figure 9 Fraction of nodes that are part of virtual forwarding backbone (RIMA and gateway nodes) as node density increases

b) Delay performance

Figure 10 shows average delay as the number of sleep cycles in a beacon periods are increased. As can be seen, delay goes on decreasing as number of sleep cycles per beacon period is increased. This is because, to deliver packet at the last hop, RIMA node has to wait for less amount of time, if number of sleep cycle per beacon period is more. It can also be seen that with load distribution delay has been reduced. For more number of sleep cycle per beacon period, average delay drops.

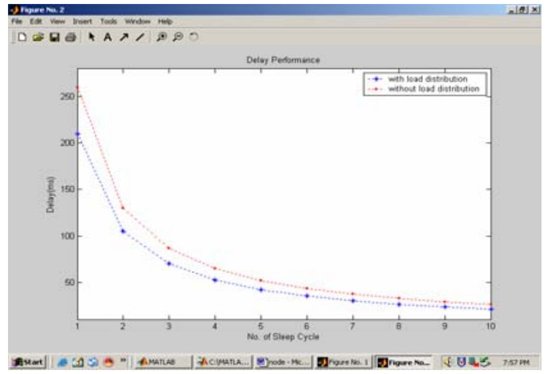


Figure 10: Average delay for CBR traffic

c) Overhead messages per node

Figure 11 shows the comparison of overhead messages of topology management scheme and routing as the number of nodes increase. Number of overhead messages per node per second decrease as number of nodes increase. Also it can been seen that overhead messages per node per second with Topology Management scheme is less as compared with flat topology.

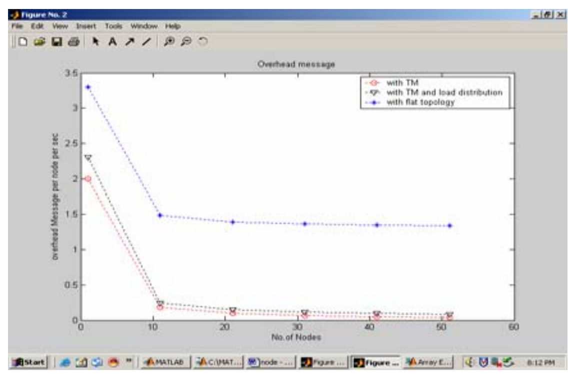


Fig 11 Overhead messages per node per second

d) Power consumption

Figure 12 shows the average power consumption, as node density increases. It can be noticed that as node density increases, average power consumption per node is much less in Topology Management scheme, as compared to flat Topology network. For more node density, there are less number of RIMA and gateway nodes, which are awake all the time and large number of member nodes are in power efficient mode, most of the time.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

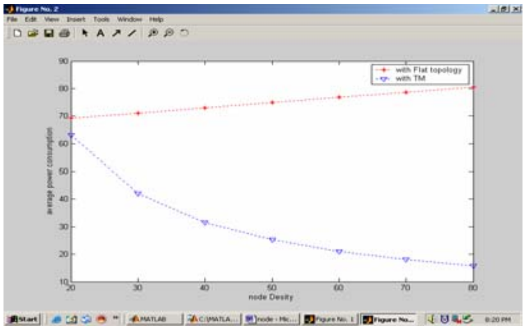


Fig 12 average power consumption as node densiy increses

e) Node lifetime

Figure 13 shows fraction of nodes remaining in the network as a function of simulation time.

If the energy of a node falls below 50 J, the node is Marked as dead. Figure 13 shows that Topology Management scheme increases the lifetime of node more than factor of two.

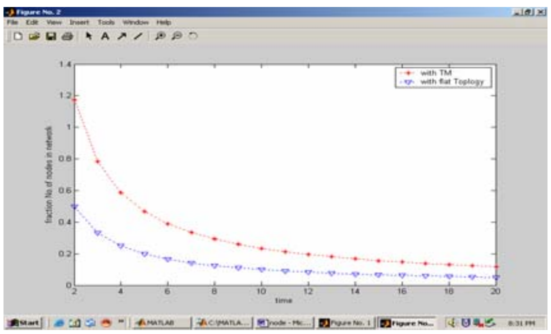


Fig 13 Node lifetime

VI. CONCLUSION

the proposed RMP protocol enable the system to operate despite the presence of misbehavior and keeps the network functional for normal nodes when other nodes don't route and forward correctly and the protocol is integrated with modified Bayesian approach to decide whether node is misbehaving or not. We conclude that DSR with RMP can give better performance than original DSR to cope with misbehavior and the of second hand information speeds up the detection of misbehaving nodes. a proposed power management scheme, which may be expected to exhibit energy saving performance, which is at least comparable to the best-published results, as well as some advantage in simplicity and extensibility is suggested.

REFERENCES

[1].A review of routing protocols for mobile ad hoc networks: by Mehran Abolhasan, Tadeusz wysocki,Eryk Dutkiewicz 2003.

[2] .MarcoCarbone, Mogens Nielsen, and Vladimiro Sassone. A formal model for trust in dynamic networks. BRICS Report RS- 03-4, 2003.

[3]. Dave B. Johnson and David A. Maltz. The dynamic source routing protocol for mobile ad hoc networks. Internet Draft, Mobile Ad Hoc Network (MANET)Working Group, IETF, October 1999.

[4] Sonja Buchegger and Jean-Yves Le Boudec. Performance Analysis of the CONFIDANT Protocol: Cooperation Of Nodes—Fairness In Dynamic Adhoc NeTworks. In Proceedings of IEEE/ACM Symposium on Mobile Ad Hoc Networking and Computing (MobiHOC), Lausanne, CH, June 2002. IEEE.

[5] Sonja Buchegger and Jean-Yves Le Boudec. A robust reputation system for mobile ad-hoc networks. EPFL Technical Report No. IC/2003/50, July 2003.

[6] Levente Butty´an and Jean-Pierre Hubaux. Enforcing service availability in mobile ad-hoc wans. In Proceedings of IEEE/ACM Workshop on Mobile Ad Hoc Networking and Computing (MobiHOC), Boston, MA, USA, August 2000.

[7] Levente Butty´an and Jean-Pierre Hubaux. Stimulating cooperation in selforganizing mobile ad hoc networks. Technical Report DSC/2001/046, EPFL-DI-ICA, August 2001.

[8] Chrysanthos Dellarocas. Immunizing online reputation reporting systems against unfair ratings and discriminatory behavior. In Proceedings of the ACM Conference on Electronic Commerce, pages 150–157, 2000.

[9] John R. Douceur. The sybil attack. In Proc. of the IPTPS02 Workshop, Cambridge, MA (USA), March 2002.

[10] Yih-Chun Hu, David B. Johnson, and Adrian Perrig. SEAD: secure efficient distance s (WMCSA 2002), IEEE, Calicoon, NY, to appear., June 2002.

[11] Yih-Chun Hu, Adrian Perrig, and David B. Johnson. Ariadne: A secure on-demand routing

[15] Sergio Marti, T.J. Giuli, Kevin Lai, and Mary Baker. Mitigating routing misbehavior in mobile ad hoc networks. In Proceedings of MOBICOM 2000, pages 255–265, 2000.

[16] Pietro Michiardi and Refik Molva. CORE: A collaborative reputation mechanism to enforce node cooperation in mobile ad hoc networks. Sixth IFIP conference on security communications, and multimedia (CMS 2002), Portoroz, Slovenia., 2002.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

PERFORMANCE OF HYBRID ROUTING PROTOCOL FOR ADHOC NETWORK UNDER BANDWIDTH CONSTRAINTS

A K Daniel
Assistant Professor
Computer Sc & Engg Department
M M M Engineering College
GORAKHPUR (U P) India
danielak@rediffmail.com

R Singh
Assistant Professor
Department of CS & I T
M J P Rohilkhand University
BAREILLY (U P) India
rsiet2002@gmail.com

J P Saini
Principal
M M M Engineering College
GORAKHPUR (U P) India
Jps_uptu@rediffmail.com

ABSTRACT: ***An Ad hoc network is a collection of wireless mobile nodes dynamically forming a temporary network without the use of any existing network infrastructure or centralized administration. Routing protocols used inside ad hoc networks must be prepared to automatically adjust to an environment that can vary between the extremes of high mobility with low band width, and low mobility with high bandwidth. In this paper, a bandwidth-efficient multicast routing protocol for ad-hoc networks is presented. A hybrid routing protocol under bandwidth constraints (HRP-BC) has been proposed. The proposed protocol achieves low communication over head, and achieves high multicast efficiency this protocol has improved existing routing protocols by creating a mesh and providing multiple alternate routes. The protocol considered the following 1) Route setup as routing Distance of path, 2) Load at the node as traffic and 3) Bandwidth as queue length at the node. The proposed scheme utilizes the path information, traffic and bandwidth resource information at each node, for selection of route path, and compared to traditional DSR schemes. The simulation results shows that the proposed HRP-BC protocol achieves better performance to the DSR protocol for the maintenance overhead and the path reliability. It reduces congestion in network and improves bandwidth utilization. Thus provide efficient use of bandwidth in the ad hoc network.***

Keywords *MANET*, Proactive, Reactive, Hybrid, bandwidth

INTRODUCTION

An ad hoc network is a collection of wireless mobile nodes dynamically forming a temporary network without the use of any exiting network infrastructures or centralized administration. Ad-hoc networks are self-configuring and self-maintaining networks that allow dispensing of fixed infrastructures. The networks rely on nodes cooperation for providing packet routing. Ad-hoc network technology presents a great potential in application domains where infrastructure deployment is expensive or not possible, like battlefield environments [1], transportation [2] or ambient intelligence scenarios [3]. Cornerstones of ad hoc networks are routing protocols. These protocols are specifically designed to promote dissemination of routing information among network nodes. The goal is to allow the creation of communication links between any two network nodes and responsible for enabling network communications. While exchanging information, the nodes may continue to move, so the network must be prepared to adapt continually [4][5][6]. The network infrastructure component such as repeaters or base-stations will frequently be either undesirable or not directly reachable, the nodes must be prepared to organize themselves into a network and establish routes among themselves without any outside support. In the simplest cases, the nodes may be able to communicate directly with each other. However, ad hoc networks must also support communication between nodes that are only indirectly connected by a series of hops through other nodes. In general, an ad hoc network looks like a network in which every mobile node is potentially a router, and all nodes run a routing protocol[7][8]. Unfortunately, standard routing algorithms work poorly in a mobile environment in which network topology changes may be drastic and frequent as the individual mobile nodes move. Such protocols are specifically designed to work in absence of fixed infrastructures. In order to promote dissemination of routing information, nodes must cooperate and rely on each other to provide routing services. To allow the creation of communication links between any two Network nodes, nodes can function both as end-hosts within each node's radio range, and as intermediate routers for other network nodes far apart[9][10]. A link is a one hop connection between two nodes. A set of links enabling the communication between a source and a destination defines as multi-hop route. The applications running inside network nodes communicate among them through data flows and. exchanging of different types of data

packets in side the network. In ad hoc mobile networks, routes are mainly multihop because of the limited radio propagation range and topology changes frequently and unpredictably since each network host moves randomly. The routing protocols can be divided into two types reactive and proactive. The Reactive routing protocols find a route on demand by flooding the network with Route Request packets. Conversely, proactive routing protocols maintain fresh lists of destinations and their routes by periodically distributing routing tables , therefore routing is an integral part of ad hoc communications, and has received interests from many researchers. Furthermore, the wireless channel is also a shared-access medium and the available bandwidth also varies with the number of hosts contending for the channel. Due to its ease of deployment and no centralized control unit, mobile nodes can connect with each other in any form of network topology at anytime[11][12]. All mobile nodes serve as routers and maintain the dynamic time-varying network topology. In MANETs, multicasting service plays an important role in bandwidth saving for some applications such as emergent events one to- one unicast transports. Without multicast capability, data Stream must be sent to all receivers by multiple unicast connections. Several researches deal with this issue recently. The network utilization will become inefficient and much more transmission and control overheads will be introduced [13]. There are many multicast protocols in traditional wired networks such as Distance Vector Multicast Routing Protocol(DVMRP [14][15][16], Multicast extension to the Open Shortest Path First , Core Based Trees (CBT) [17], Protocol Independent Multicast,. Core Based Tree (CBR) is a tree based multicast protocol.[18] The main idea of this protocol is to find a nearest forwarding node to replace with finding the shortest path between the source node and the receiving node for decreasing the number of packet transmission. Multicast Ad hoc On-Demand Distance Vector Routing Protocol (MAODV) [20] is a modified version of AODV [19]. In MAODV, each node of MANET must send control packets periodically to maintain the topology. Weight-Based Multicast Protocol is also a tree based multicast protocol Bandwidth-Efficient Multicast Routing Protocol (BEMRP) [21]. However, it not only considers the transmission hop but also considers the overhead of the forwarding path. Generally, the above schemes consider the multi path connection at routing layer and leave the issue of reliable transmission being dealt with at upper layer.[22][23] Dynamic change in topology with time in MANET gives ,several issues, such as processing overhead, packets collisions, and route maintaining, need to be overcome Network population and node density are important concerns for ad-hoc networking; the higher the population and node density, the higher the probability to reach any network node [24][25] [26]. Ad hoc networks are thus conceived to ease the entrance of new nodes. This property is their biggest strength as well as their main security weakness. In this paper, the proposed HRP-BM protocol deals with the issue of reliable multicast to reduce the maintenance overhead increase the path stability, reducing the congestion in mobile ad-hoc network and efficient use of bandwidth. The proposed protocol requires only a small number of control packets to setup and maintain multicast routes as well as a small number of packet transmissions to deliver multicast packets to the receivers, and has high multicast efficiency with low communication overhead.

The rest of the paper is organized as follows. The problem statement is given in section II .The proposed model and algorithm to solve the problem is given in section III. The comparative results are discussed in section IV Finally, conclusions and future work are discussed in Section V and VI respectively.

II.PROBLEM STATEMENT: ROUTING AND BAND WIDTH MANAGEMENT IN ADHOC NETWORKS

Ad hoc network has emerged as one of the most focused research areas in the field of wireless networks and mobile computing. Ad hoc networks consist of hosts communicating one another with portable radios The basic routing problem is that of finding an ordered series of intermediate nodes that can transport a packet across a network from its source to its destination by forwarding the packet along the series of intermediate nodes. In traditional hop-by-hop solutions to the routing problem, each node in the network maintains a routing table. For each known destination, the routing table lists the next node to which a packet for the destination should be sent. The routing table at each node can be thought of as part of a distributed data structure that, taken together, represents the topology of the network. The goal of the routing protocol is to ensure that the overall data structure contains a consistent and correct view of the actual network topology. If the routing tables at some nodes were to become inconsistent, then packets can loop in the network. If the routing tables were to contain incorrect information, then packets can be dropped. The problem of maintaining a consistent and correct view becomes harder as there is an increase in the number of nodes whose information must be consistent, and as the rate of change in the true topology increases. The challenge in creating a routing protocol for Ad hoc networks is to design a single protocol that can adapt to the wide variety of conditions present inside ad hoc networks. For example, the bandwidth available between two nodes in the network may vary from more than 10 Mbps, when using high-speed network interfaces with little interference, to 10 Kbps or less when using low-speed network interfaces or when there is significant interference from outside sources or other nodes'

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

transmitters. Similarly, nodes in an Ad hoc network may alternate between periods when they are stationary with respect to each other and periods when change topology, rapidly conditions across a single network may also vary, so that some nodes are slow moving, while others change location rapidly. The routing protocol must perform efficiently in environments in which nodes are stationary and bandwidth is not a limiting factor. Yet, the same protocol must still function efficiently when the bandwidth available between nodes is low and the level of mobility and topology change high. Because it is often impossible to know a priori what environment the protocol will find itself in, and the environment can change unpredictably, the routing protocol must be able to adapt automatically. Most routing protocols include at least some periodic behaviors, meaning that there are protocol operations that are performed regularly at some interval regardless of outside events. These periodic behaviors typically limit the ability of the protocols to adapt to changing environments. If the periodic interval is set too short, the protocol will be inefficient as it performs its activities more often than required to react to changes in the network topology. If the periodic interval is set too long, the protocol will not react sufficiently to changes in the network topology quickly and lost packets Periodic protocols can be designed to adjust their periodic interval to try to match the rate of change in the network, but this approach will suffer from the overhead associated with the tuning mechanism and the lag between a change in conditions and the selection of a new periodic interval. In the worst case, which consists of bursts of topology change followed by stable periods, adapting the periodic interval could result in the protocol using a long interval during the burst periods and a short interval in the stable periods. This worst case may be fairly common, for example, as when a group of people enter a room for a meeting, are seated for the course of the meeting, and then stand up to leave at the end. The alternative to a periodic routing protocol is one that operates in an *on-demand* fashion. On-demand protocols are based on the premise that if a problem or inconsistent state can be detected before it causes permanent harm, then all work to correct a problem or maintain consistent state can be delayed until it is proven to be needed. They operate using the same “lazy” philosophy as optimistic algorithms. The Dynamic Source Routing protocol (DSR) is unique among the current set of routing protocols for ad hoc networks in the way it avoids periodic behavior, and in the way it solves the routing information consistency problem. First, DSR is completely on-demand, which causes the overhead of the protocol to automatically scale directly with the need for reaction to topology change. This dramatically lowers the overhead of the protocol by eliminating the need for any periodic activities, such as the route advertisement and neighbor detection packets that are present in other protocols. Second, DSR uses source routes to control the forwarding of packets through the network. The key advantage of a source routing design is that intermediate nodes do not need to maintain consistent global routing information, since the packets themselves already contain all the routing decisions. Beyond this, the source route on each packet describes a path through the network. Therefore, with a cost of no additional packets, every node overhearing a source route learns a way to reach all nodes listed on the route [27] [28].

III DESIGN SPACE AND PROPOSED ALGORITHM

The proposed multicast routing protocol requires low Communication overhead since it does not require periodical transmission of control packets. Most of the existing multicast routing protocols, such as DVMRP (Distance-Vector Multicast Routing Protocol) [8] and FGMP (Forwarding Group Multicast Protocol) [9], require periodical transmission of control packets in order to maintain multicast group membership and multicast routes, thereby wasting a lot of bandwidth. In the proposed protocol, route setup and route recovery are invoked only when they are required route setup process is invoked only when a new node joins a multicast group, and route recovery process is invoked only when a multicast route breaks due to the node movements. Further, in the route recovery process, control packets used to recover multicast routes are flooded only to limited network area scoped by TTL (time-to-live). In our protocol, bandwidth level at a node is used as TTL .Limiting the scope of route search further decreases the communication overhead since control packets are not flooded to the entire network but only to just previous node (predecessor node)MAODV (Multicast Ad-hoc On Demand Distance Vector) also tries to minimize the communication overhead by invoking the route discovery process on-demand. However, unlike the proposed protocol, MAODV ignores multicast efficiency.

The proposed multicast routing protocol also achieves high multicast efficiency, i.e., it requires a small number of multicast transmissions. Multicast transmission is kept minimal by keeping the number of forwarding nodes small. Forwarding nodes are the nodes which broadcasts (forwards) multicast packets to neighboring nodes. Most of the existing multicast Routing protocols use unicast protocols such as DSDV (Destination Sequenced Distance Vector) and AODV (Ad hoc On Demand Distance Vector) to select the shortest paths from a source to each receiver. For example, in CBT (Core Based Tree) and PIM (Protocol Independent Multicast) based protocols when a new node needs to join a multicast group, these unicast protocols are used to set up the shortest path to a core . In FGMP forwarding nodes are selected along the shortest paths chosen by these unicast protocols. In multicast

environment, using the shortest paths from a source to each receiver does not always result in efficient multicast. Unlike these existing multicast protocols, the proposed protocol does try to find a shortest path, instead, it tries to find the nearest forwarding node in the multicast group when a node wants to join the group. Nodes along the path between the nearest forwarding node and the new node become new forwarding nodes. This results that the minimum number of forwarding nodes are added. In addition, the proposed protocol provides a mechanism to detect unnecessary forwarding nodes and delete them from a multicast group. Due to the dynamic nature of ad-hoc environment, there may be unnecessary forwarding nodes in a multicast group. Route optimization process employed in the proposed protocol can detect and delete them from a multicast group to reduce unnecessary transmissions of multicast packets. This further increases multicast efficiency. Hybrid routing protocols combines the advantages of proactive and of reactive routing. The routing is initially established with some proactively prospected routes and then serves the demand from additionally activated nodes through reactive flooding.

PROPOSED MODEL

Mobile Ad-hoc Network (MANET) has with available bandwidth B and number of nodes be n and distance between nodes is D and load at each node be L.
The following figure shows the wireless network of five nodes as Base Station,
B= Total Available Bandwidth
n_i=Nodes Name(Base Station)
Q_i=Length of queue at node n_i
M_i=Total number of users at node n_i (Base Station)

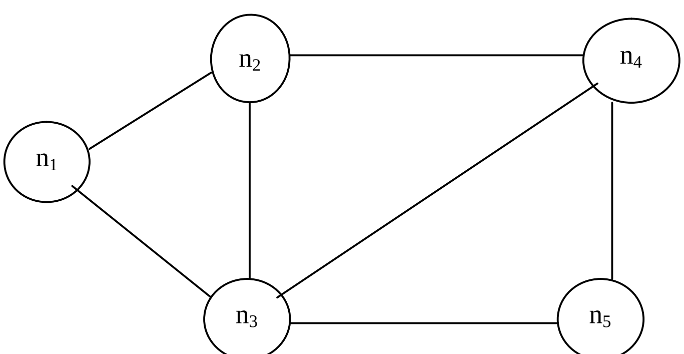


Fig 1

So in order to select path from n_1 to n_5

1. P_1:n_1-n_2-n_4 –n_5, or
2. P_2:n_1-n_3-n_5, or
3. P_3:n_1-n_2-n_3-n_4-n_5, or
4. P_4:n_1-n_2-n_3-n_5, or
5. P_5:n_1-n_2-n_4-n_3–n_5 or
6. P_6:n_1-n_3-n_4-n_5,

III (A) DESIRED CHARACTERISTIC FOR PATH SELECTION

1. Distance of selected path is minimum or optimum.
2. Load in selected path is minimum or optimum and load at intermediate node is less than threshold of B.
3. Queue length at intermediate nodes of the path is minimum or optimal.

Distance is based in the number of hop counts.
Queue Length is known to all the nodes, and while transferring the queue length, the maximum of all queue length at intermediate nodes in path is stored only. As a node can transfer only one packet at time, thus the queue length can be used to estimate the available bandwidth, as we are not considering multiplexing of data at the nodes. Thus, the position of paths in bandwidth list will be similar to the position of path in queue length list.

III (B) CONTROL PACKET DETAILS:

It have the following tables and functions

(I) ROUTING TABLE CONSTRUCTION:

Whenever a mobile node enters a wireless network it would broadcast a notification packet with fields as shown in fig. – 2

Node No.	Distance	Queue length	Flag (00)

Fig. – 2 Notification packet

Initially the distance field value is initialized to 1 and queue length is initialized to 0 and node number is calculated from the IP address and subnet mask. Arithmetic to calculate the node number is to apply AND operation on complement of subnet mask and IP address. Flag field is a 2-bit field and its set to 00 for notification packet. The receiving node, would match node no. of the received packet from their table, if it don't have this node no. registered in its table, it would add a row. An example of table construction is shown in Fig. – 3

(II) PATH CONSTRUCTION

Now when a nodes get a packet to transmit to some other node then it calculate the node number of the destination node, if it has entry of this node in its routing table then it would simply send packet to it otherwise it would broadcast the route request(RREQ) packet(Fig. – 4) with unique sequence no., its node no. as the sender and route source node no. the flag is set as 01 for RREQ. The node receiving RREQ would then check its table for destination node, if it has entry then it sends the

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

route reply packet (RREP) (Fig. – 5) packet with its sequence no. as that of RREQ packet, it's node number in the path field, distance increased by one from the value in its routing table in the distance field, it's queue length in the queue length field and flag as 10 to the sender node number of RREQ packet, and then multicast the RREQ packet with changed sequence no and its node no. as sender node no. to all the nodes in its routing table except of sender node no. and destination node no and save this information of original sequence no., modified sequence no. sender node no. and route source node no. in its memory. Even if the node receiving RREQ does not have the entry of destination node in its routing table then also it would multicast it to other node, in its table, except for sender node, route source node number and destination node no in the same manner as described before. The node receiving route reply packet checks to see if it itself is route source node number. If it is not the route source node no. then it would match the sequence no. and route source node no. in its memory to retrieve the original sequence no. and sender node no. and it sends the route reply packet to the sender node no. with sequence no. as

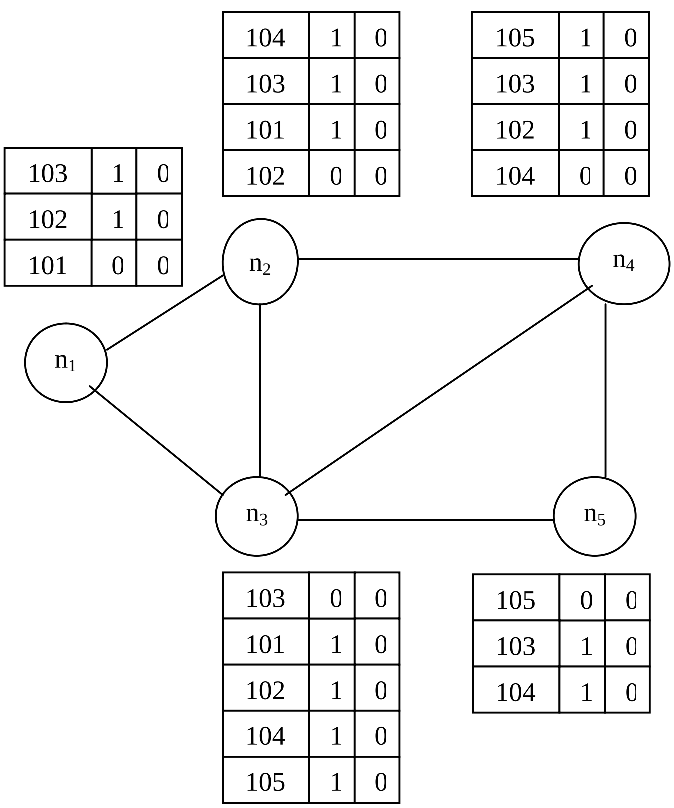


Fig. – 3 Construction of Routing Table

Sequence No.	Path data	Route Source Node No.	Destination Node No.	Dist-ance	Queue Length	Flag (10)

Fig. – 4 Route Request Packet (RREQ)

Path data	Destination Node No.	Route Source Node No.	Data	Flag (11)

Fig. – 5 Route Reply Packet (RREP)

Sequence No.	Sender Node No.	Route Source Node No.	Destination Node No.	Flag (01)

Fig. – 6 Message Packet

retrieved sequence no. and append its own node no. to the path data field and queue field value is set either to value of queue length field of received RREP packet or the its queue length value from the routing table depending upon whichever is maximum.

The route source node on receiving any of the, RREP packet will save it in its path list as shown in Fig. – 7.

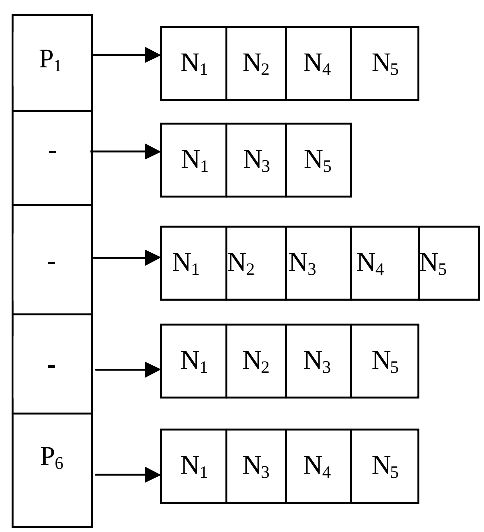


Fig – 7 Path list at the Route Source Node

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

(III) ALGORITHM: ROUTES MAIN-TENANCE

Input : Routing Table: RTable [] [], MessagePacket :M[], Destination Node No. : D_node , Boolean variable Flag=0

1. Start
2. Len=Length[M]
3. If ((M [Len – 2] = = 1) AND (M [Len – 1] = = 1))/*Message packet received*/
4. For I = 0 to Length [RTable]
5. If (RTable [i] [0] = = D_node)
6. Transmit M to D_node
7. Flag = 1
8. Break
9. End If
10. End For
11. If (! Flag)
12. Broadcast RREQ packet with field values as
    Seq (Sequence NO.) = System generated no.
    S_No.(Sender node_no.) = self node no.
    Rs No. (Route Source Node no.) = Self Node
    D_no. (Destination Node No.) = D_node
    F (Flag) = 0 1
13. End If
14. If ((M [Len – 2] = = 0) AND (M [Len – 1] = = 1)))/*Route request packet received*/
15. For I = 0 to Length [RTable]
16. If (RTable [i] [0] = = D_node)
17. Send RREP to S_No. with Field values
    Seq (Sequence No) = RREQ.Seq
    Pd (Path Data) = stack implementation (with self node no on top)
    Rs No. (Route Source Node no.) = RREQ.Rs No.
    D_no. (Destination Node No.) = RREQ . DNo.
    D (Distance) = RTable [i] [1] +1
    Q_Len (Queue Length) = RTable [i] [2]
    F (Flag) = 1 0
18. Flag = 1
19. Break
20. End If
21. End For
22. If (! Flag)
23. Multicast RREQ packet to all except for sender node no , route source node no and destination node no with field values
    Seq (Sequence NO.) = System generated no.
    S_No.(Sender node_no.) = self node no.
    Rs No. (Route Source Node no.) = Self Node
    D_no. (Destination Node No.) = D_node
    F (Flag) = 0 1
24. Make an entry in system database with field values
    New_Seq = Seq in step 23
    Old_Seq = RREQ.Seq
    RS_No. = RREQ.RS_No.
    D_No. = RREQ.D>No.
    Sender = RREQ.SNo.
25. End If
26. If ((M [Len – 2] = = 1) AND (M [Len – 1] = = 0)) /*Route reply packet received*/
27. If (RREP.RS_No. = = Node_No.)
28. Add Path data of RREP to the path Linked List at the node.
29. Else
30. Insert its node no. in path data of RREP
31. If (RREP.Q_Len < RTable[0][2])
32. RREP.Q_Len=RTable[0][2]
33. End If
34. Retrieve sender node no. and Sequence number from database by RREP.Seq, RREP.S_no.
35. RREP.Seq=Sequence no. of step 33
36. Send RREP to sender node of sep 33
37. End If
38. End If
39. Stop.

(IV) ALGORITHM: PATH SELECTION

Consider the following paths selection rules

- Arrange all the possible paths in ascending order of queue length, load and distance, consider only paths which has load lower than threshold.
- Take the sum of position of the path in the three lists and finally select the path with lowest sum.
- In case if minimum sum of position in the three lists calculated in step (ii) is more than one then the following preference order is used for selection of an optimal path.

Queue Length > Load > Distance of path

The queue length (Bandwidth Concept) of each node in the fig-1 is as follows
Q_1=10, Q_2=12, Q_3=15, Q_4=9, Q_5=5

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Thus the queue length and distance of paths are shown in table-1 as details of Paths.

Distance	Path	Queue Length
3	P_1 :n_1-n_2-n_4-n_5	10
2	P_2 :n_1-n_3-n_5	15
4	P_3 :n_1-n_2-n_3-n_4-	15
3	P_4 :n_1-n_2-n_3-n_5	15
3	P5 n1-n3-n4-n5	15
4	P6 :n1-n2-n4-n3-n5	15

Table 1

Arranging the paths in ascending order with respect to Distance, load and queue length as follows in the table with their position Table 2

Position	Distance	Load	Queue Length
1	P2	P1	P1
2	P1	P2	P2
3	P4	P3	P3
4	P3	P4	P4
5	P5	P5	P5
6	P6	P6	P6

Table 2

The sum of position of path in the three lists (distance, load and queue length)

For p_1: (2+1+1) =4
For p_2: (1+2+2) =5
For p_3: (4+3+3) =10
For p_4: (3+4+4) =11
For p_5: (5+5+5) =15
For p_6: (6+6+6) =18

From above calculation it is clear that the sum of position of path P_1 in the three lists is minimum hence path P_1 is selected.

IV SIMULATION RESULTS

In simulation, a flat network is assumed as clusters. For unicast, before a node sends a unicast packet, it sets RTS (Request-to- Send) flags of its neighbors and the intended receiver sets CTS (Clear-to-Send) flags of its neighbors. Nodes whose RTS or CTS flag is set cannot transmit data, except the sender. When the sender finishes sending the data, RTS/CTS flags are cleared by the nodes which originally set those flags. Similar scheme is used in multicasting .The node wants to send a multicast packet sets RTS flags of its neighbors, and each intended receiver sets CTS flags of its neighbors. The broadcast uses flooding, technique and only RTS flags are set by the sending node, and CTS flags are not set by any node. Therefore, in broadcast, collision may occur. However, collisions are ignored in our simulation. The simulated network area is a N x N meter square, and M mobile nodes are roaming randomly in all directions at a predefined speed in this area. Each node has a finite buffer, and packets are lost when buffer overflow occurs. Control packets have higher priority over data packets in simulations1. Propagation delay is assumed to be negligible, and it is assumed that packets always arrive without any bit error. A multicast group has one source and a number of receivers. The source node generates multicast packets at a constant rate Extensive simulation results obtained by varying several network parameters and workload configuration. The values of the network parameters used in simulations are those specified in the IEEE 802.11. In this scenario we evaluate the performance improvement in terms throughput due to the use of multiple channels in a densely populated network. Specifically, we consider a network of 5 to 20 Base Stations with 20 to 80 mobile nodes with an increasing number of neighbors from 5 to 20 Base Station. Each node has a traffic flow with infinite demands towards one of its neighbors. In Fig. 9 to Fig. 13.We show the some of throughput of all traffic flows, with available Channel Bandwidth.

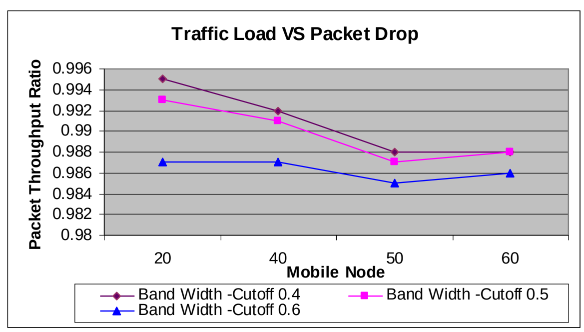


Fig-9

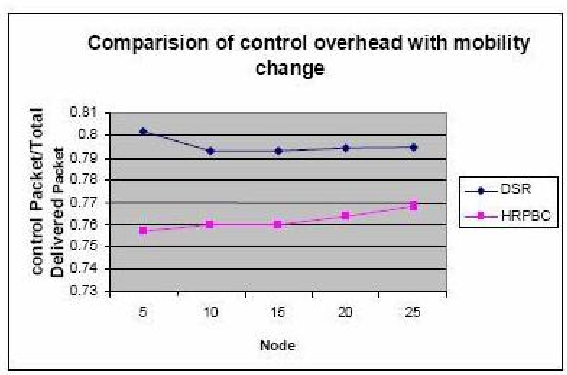


Fig. – 10

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

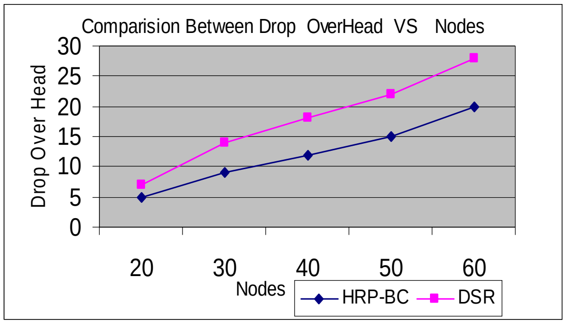


Fig-11

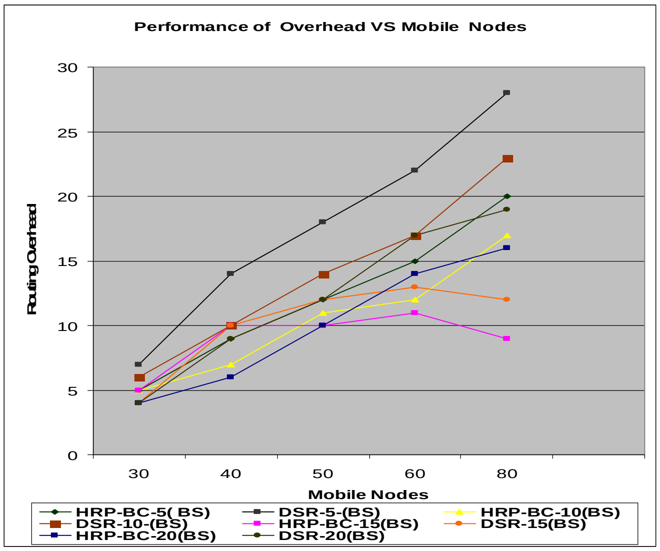


Fig-12

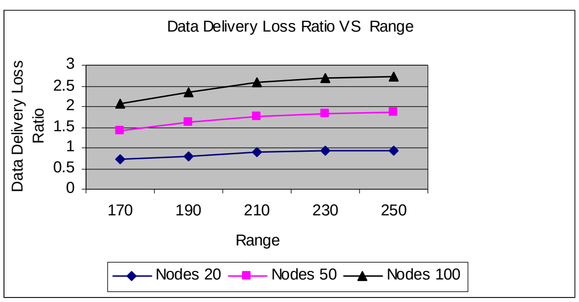


Fig-13

V CONCLUSION

The proposed HRPBC protocols will mostly select the optimal path for transmission of packets from source to destination in wireless Ad-hoc networks and adopts the path information kept at each node with bandwidth information. It is compared to traditional DSR schemes. The simulation show that the proposed HRPBC protocol achieve the above objectives and is superior to that of the DSR scheme for the maintenance overhead and the path reliability, Thus reducing the congestion in network and improving bandwidth utilization.

VI FUTURE SCOPE

In future there can be further evaluation of proposed scheme by using more realistic mobility of nodes in the simulation. It is believed that advantage of providing traffic information will be significant in those environments.

VII REFERENCES

[1] D. B. Johnson and D. A. Maltz. Dynamic source routing in ad-hoc wireless networks. Mobile Computing Kluwer Academic Publishers, 1996.

[2] C-K Toh. Wireless ATM and Ad-Hoc Networks: Protocols and Architectures. Kluwer Academic Publishers, 1997

[3] C. E. Perkins and P. Bhagwat. Highly Dynamic Destination Sequenced Distance Vector Routing (DSDV) for Mobile Computers, SIGCOMM Conf. Proc, 1994.

[4] S. Corson and A. Emphremides. A Distributed Routing Algorithm for Mobile Wireless Networks, ACM/Baltzer Wireless Networks J., vol. 1, no.1, 1995.

[5] S. Murthy and J. J. Garcia-Luna-Aceves. A Routing Protocol for Packet Radio Networks, MOBICOM, 1995

[6] R. Dube et. Al. Signal Stability-Based Adaptive Routing (SSA) for Ad Hoc Mobile Network, IEEE Personal Communications, 1997

[7] Z. J. Hass. A New Routing Protocol for the Reconfigurable Wireless Network, ICUPC, 1997

[8] S. E. Deering and D. R. Cheriton. Multicast Routing in Datagram Internetworks andExtended LANs. ACM Transaction on Computer Systems, May 1990.

[9] C-C. Chiang and M. Gerla. On-Demand Multicast in Mobile Wireless Networks Proceedings of IEEE ICNP '98, 1998

[10] C-C. Chiang and M. Gerla. Routing and Multicast in Multihop, Mobile Wireless Networks Proceedings of ICUPC '97, 1997

[11] C-C. Chiang, M. Gerla and L Zhang. Shared Tree Wireless Network Multicast. Proceedings of IEEE 6th International Conference on Computer Communications and Networks (ICCCN'97), 1997.

[12]C. E. Perkins. Ad-hoc On Demand Distance Vector (AODV) routing Proceedings of the IEEE 1997

[13] David B. Johnson and David A. Maltz. Dynamic source routing in ad hoc wireless networks Technical report, Carnegie Mellon Unity, 1996

[14] D. Waitzman, C. Partridge, and S. Deering, “Distance Vector Multicast Routing Protocol (DVMRP)”, RFC 1075, Nov. 1988

[15] B. Quinn, and K. Almeroth, “IP Multicast Applications: Challenges and Solutions”, RFC 3170, Sep. 2001

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[16] C. E. Perkins and P. Bhagwat. Highly dynamic Destination- Sequenced Distance-Vector routing (DSDV) for mobile computers. Proceedings of the SIGCOMM '94, page 234-244, August 1994.

[17] T. Ballardie, P. Francis, and J. Crowcroft. Core Based Tree (CBT) an architecture for scalable interdomain multicast routing. Proceeding of ACM SIGCOM, 1993

[18] A. Ballardie, "Core Based Trees (CBT version 2) Multicast Routing Protocol Specification", RFC 2186, Sep. 1997.

[19] C. PERKINS, E.ROYER AND S. DAS Ad hoc On-demand Distance Vector (AODV) Routing, RFC

[20] E. M. Royer and C. E. Perkins. Multicast Ad hoc On-Demand Distance Vector (MAODV) Routing . draftietf. manet-maodv-00.txt, July 2000.

[21] T. Ozaki, J. B. Kim, and T. Suda, "Bandwidth-Efficient Multicast Routing for Multihop, Ad-Hoc Wireless Networks", in IEEE INFOCOM, 2001.

[22] E. M. Royer, and Charles E. Perkins," Multicast Operation of the Ad hoc On-Demand Distance Vector Routing Protocol", Proceedings of ACM MOBICOM 1999, pp. 207-218

[23] Zbigniew Dziong, Marek Juda, and Lorne G. Mason.] " A Framework for Bandwidth Management in ATM Networks -Aggregate Equivalent Bandwidth Estimation Approach" pp 134-146

[24] Ad-hoc On-demand Multipath Distance Vector - M. Marina, S. Das: On-demand Multipath Distance Vector Routing in Ad Hoc Networks, Proceedings of the 2001 IEEE International Conference on Network Protocols (ICNP) IEEE Computer Society Press, 2001.

[25] Guangyu Pei and Mario Gerla and Xiaoyan Hong Ching-Chuan Chiang, A Wireless Hierarchical Routing Protocol with Group Mobility, IEEE WCNC'99, New Orleans, USA, September 1999.

[26] C. E. PERKINS, P. BHAGWAT Highly Dynamic Destination-Sequenced Distance Vector (DSDV) for Mobile Computers Proc. of the SIGCOMM 1994 Conference on Communications Architectures, Protocols and Applications, Aug 1994,.

[27]Internet Engineering Task Force (IETF) Mobile Ad Hoc Networks (MANET) Working Group Charter, Chaired by Joseph Macker and M. Scott Corson

[28] J. Jubin and J.D. Tornow, "The DARPA Packet Radio Network Protocols," Proceedings of the IEEE, 2007

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

MVDR an Optimum Beamformer for a Smart Antenna System in CDMA Environment

M Yasin[1], Pervez Akhtar[2], M Junaid Khan[3]

Department of Electronics and Power Engineering
[1, 2, 3]Pakistan Navy Engineering College, NUST, Karachi, PAKISTAN
myasin@pnec.edu.pk, pervez@pnec.edu.pk, contactjunaid@yahoo.com

Abstract: **Efficient utilization of limited radio frequency spectrum is only possible to use smart/adaptive antenna array system. Minimum Variance Distortionless Response (MVDR) algorithm is an option for smart antenna to exploit spatial distribution of the users and the access delay distribution of signal paths to enhance mobile systems capabilities for quality voice and data communication. This paper analyzes the performance of MVDR (blind algorithm) and Kernel Affine Projection Algorithm (KAPA) (nonblind algorithm) for CDMA application. For the first time, KAPA is implemented in [1] in the context of noise cancellation but we are using it for adaptive beamforming which is novel in this application. Smart antenna incorporates these algorithms in coded form which calculates optimum weight vector which minimizes the total received power except the power coming from desired direction. Simulation results verify that MVDR a blind algorithm has high resolution not only for beam formation but also better for null generation as compared to nonblind algorithm KAPA. Therefore, MVDR is found more efficient Beamformer.**

Keywords: Adaptive Filtering, Minimum Variance Distortionless Response (MVDR) Algorithm and Kernel Affine Projection Algorithm (KAPA).

I. INTRODUCTION

Since Radio Frequency (RF) spectrum is limited and its efficient use is only possible by employing smart/adaptive antenna array system to exploit spatial distribution of the users and the access delay distribution of signal paths to enhance mobile systems capabilities for data and voice communication. The name smart refers to the signal processing capability that forms vital part of the smart/adaptive antenna system which controls the antenna pattern by updating a set of antenna weights. Smart antenna, supported by signal processing capability, points narrow beam towards desired users but at the same time introduces null towards interferers, thus improving the performance of mobile communication systems in terms of channel capacity, extending range coverage, tailoring beam shape and steering multiple beams to track many mobiles electronically. Consider a smart antenna system with Ne elements equally spaced (d) and user's signal arrives from an angle Φ_0 as shown in Fig 1 [2].

.

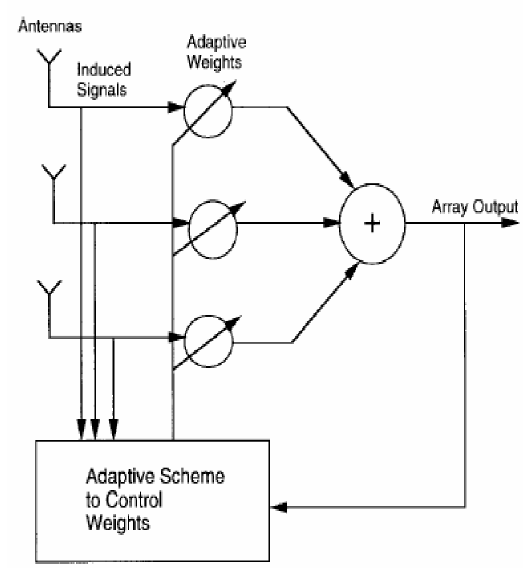


Fig.1. Smart/adaptive antenna array system

Adaptive beamforming scheme that is MVDR (blind algorithm) and KAPA (nonblind algorithm) is used to control weights adaptively to optimize signal to noise ratio (SNR) of the desired signal in look direction Φ_0. The array factor for (Ne) elements equally spaced (d) linear array is given by

$$AF(\Phi) = \sum_{n=0}^{N-1} A_n . e^{(jn(\frac{2\pi d}{\lambda}\cos\Phi + \alpha))} \qquad (1)$$

where α is the inter element phase shift and is described as:

$$\alpha = \frac{-2\pi d}{\lambda_0}\cos\Phi_0 \qquad (2)$$

and Φ_0 is the desired direction of the beam.

In reality antennas are not smart; it is the digital signal processing, along with the antenna, which makes

the system smart. When smart antenna is deployed in mobile communication in Code Division Multiple Access (CDMA) environment in which different codes are assigned to different users, it radiates beam towards desired users only. Each beam becomes a channel, thus avoiding interference in a cell. Because of these, each coded channel reduces co-channel interference, due to the processing gain of the system. The processing gain (PG) of the CDMA system is described as:

$$PG = 10\log(B / R_b) \tag{3}$$

where B is the CDMA channel bandwidth and R_b is the information rate in bits per second.

If a single antenna is used for CDMA system, then this system supports a maximum of 31 users. When an array of five elements is employed instead of single antenna, then capacity of CDMA system can be increased more than four times. It can be further enhanced if array of more elements are used [4] [5] [7] [8] [9].

The rest of the paper is organized as follows: Section 2 introduces MVDR algorithm with simulation results. KAPA with simulation results are presented in section 3. Finally the concluding remarks of this work are provided in section 4.

II. MVDR ALGORITHM

A. *Theory*

MVDR is a direction of arrival (DOA) estimation method in which the direction of a target signal is parameterized by the variable Φ_0 and all other sources are considered as interferences. In beamforming literature, this estimation method is called MVDR in which the weights of the smart antenna array are chosen so as to pass the desired directional signal without any distortion (preserving unity gain) whereas to suppress the interferers maximally. MVDR is a blind algorithm which doesn't require a training signal to update its complex weights vector but utilizes some of the known properties of the desired signal. Assuming that $s(\Phi_0)$ is the steering vector and is independent of the data obtained from n sensors. The data obtained from n sensors is given by

$$u(n) = \{u_0, u_1, ...u_{n-1}\} \tag{4}$$

MVDR beamformer output $y(n)$ in the look direction with input signal $u(n)$ is described as:

$$y(n) = w^T(n-1)u(n) \tag{5}$$

The autocorrelation matrix R of the sensor data vector is given by

$$R = E\{u(n)u^T(n)\} \tag{6}$$

where E is the expectation operator. The output power for each looking direction is defined by

$$P = E\{|y|^2\} = w^T E\{u(n)u(n)^T\}w = w^T R w \tag{7}$$

In adaptive beamforming algorithm, the weight vectors are correlated with the incoming data so as to optimize the weight vectors for high resolution DOA detection in a noisy environment. MVDR is graded an adaptive beamformer, therefore, some constraints are imposed as (8), ensures that desired signals are passed with unity gain from looking direction whereas the output power contributed by interfering signals from all other directions are minimized using a minimization criterion as described in (9).

$$w^T s = g \tag{8}$$

where g denotes the gain of MVDR which is equal to unity.

$$\underset{w}{Min}(P = w^T R w) \text{ constrained to } w^T s = 1 \tag{9}$$

Solving (9) by Lagrange multiplier method, we obtain the weight vector as:

$$w = \frac{R^{-1}s}{s^T R^{-1} s} \tag{10}$$

When we put the value of (10) into (9), the output power $P(\Phi_0)$ for a single looking direction is obtained as:

$$P(\Phi_0) = \frac{1}{s^T R^{-1} s} \tag{11}$$

MVDR algorithm computes the optimum weight vector based on the sampled data that ultimately forms a beampattern and places null towards interferers [3] [6].

B. *Simulation Results*

Computer simulation is carried out, to illustrate that how various parameters such as number of elements (Ne)

.

and element spacing (d), affect the beam formation. The simulations are designed to analyze the properties of MVDR and KAPA algorithms. The desired signal is phase modulated, used for simulation purpose. It is given by

$$S(t) = e^{j\sin(2*\pi*f*t)} \qquad (12)$$

where f is the frequency in Hertz.

1) Effect of number of elements on array factor

Uniform linear array is taken with different number of elements for simulation purpose. The spacing between array elements is taken as $(\lambda/8)$.

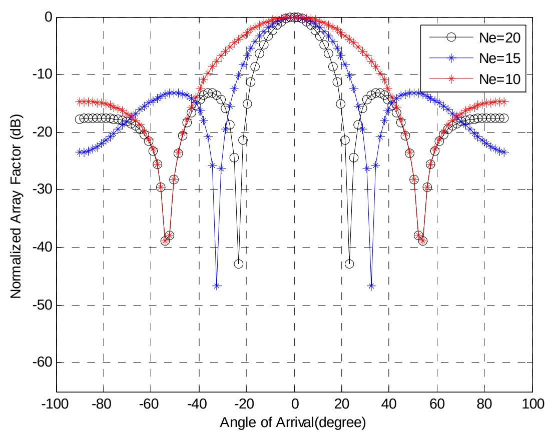


Fig.2. Normalized array factor plot for MVDR algorithm with AOA for desired user is 0 degree and **- 30 degrees for interferer** with constant space of $(\lambda/8)$ between elements

Angle of Arrival (AOA) for desired user is set at 0 degree and for interferer at -30 degrees as shown in Fig. 2 which provides deep null at -30 degrees but at the same time forms narrow beam in accordance to number of elements.

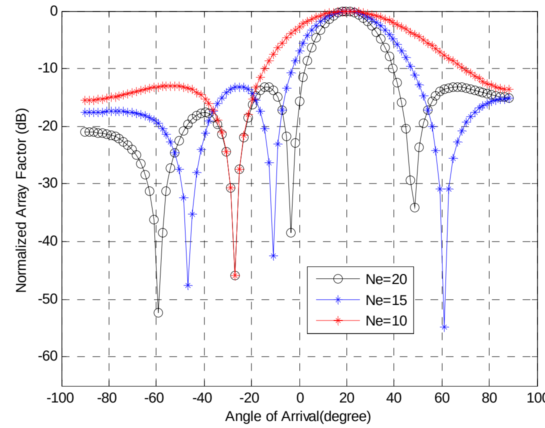


Fig.3. Normalized array factor plot for MVDR algorithm with AOA for desired user is 20 degrees and **- 20 degrees for interferer** with constant space of $(\lambda/8)$ between elements

Similarly in Fig.3, we achieved a deep null approximately at -20 degrees and the desired user is arriving at 20 degrees. Therefore, it is proved that for a fixed spacing and a frequency, a longer array $(Ne = 20)$ results a narrower beam width but this happens at the cost of large number of sidelobes.

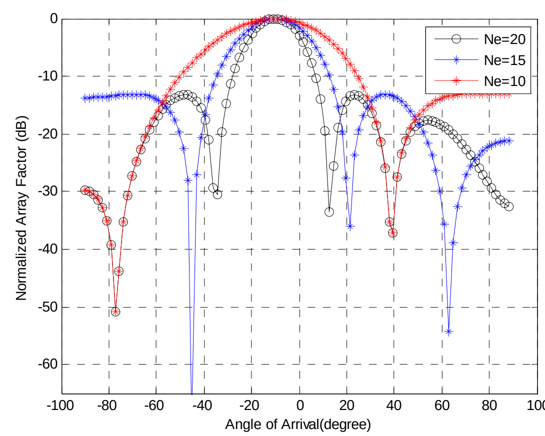


Fig.4. Normalized array factor plot for MVDR algorithm with AOA for desired user is - 10 degrees and **40 degrees for interferer** with constant space of $(\lambda/8)$ between elements

In Fig. 4, AOA for desired user is obtained at -10 degrees and deep null is shown at – 40 degrees for $d = \lambda/4$. Again it is proved that for a fixed spacing and a frequency, a longer array $(Ne = 20)$ results a narrower beam width but this happens at the cost of large number of sidelobes.

The weight vectors computed during simulation for $Ne = 20, 15$ and 10 are $w1, w2$ and $w3$, respectively as shown in Fig. 5. Numerically, these weight vectors are represented as:

$w1$ =[0.0500, 0.0482 - 0.0133i, 0.0430 - 0.0256i, 0.0346 - 0.0361i, 0.0238 - 0.0440i, 0.0113 - 0.0487i, -0.0020 - 0.0500i, -0.0152 - 0.0476i, -0.0273 - 0.0419i, -0.0375 - 0.0331i, -0.0449 - 0.0220i, -0.0491 - 0.0093i, -0.0498 + 0.0041i, -0.0470 + 0.0172i, -0.0407 + 0.0290i, -0.0316 + 0.0388i, -0.0201 + 0.0458i, -0.0073 + 0.0495i, 0.0061 + 0.0496i, 0.0191 + 0.0462i]

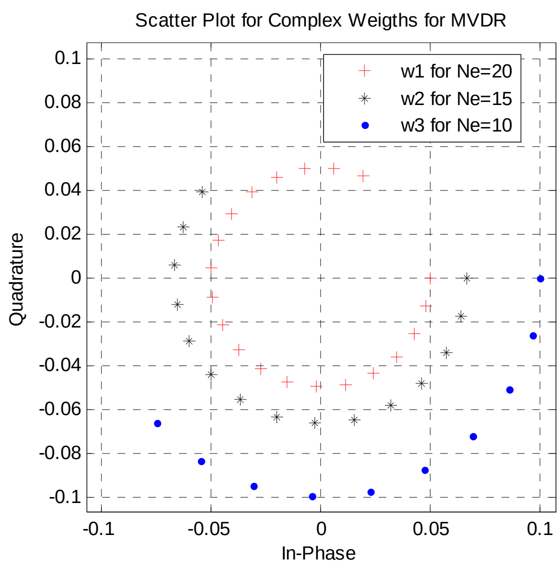


Fig.5. Scatter plot for **complex weights** for $Ne = 20, 15$ and 10 with constant space of $(\lambda / 8)$ between elements

$w2$ =[0.0667, 0.0643 - 0.0177i, 0.0573 - 0.0341i, 0.0462 - 0.0481i, 0.0317 - 0.0586i, 0.0150 - 0.0649i, -0.0027 - 0.0666i, -0.0203 - 0.0635i, -0.0364 - 0.0558i, -0.0499 - 0.0442i, -0.0599 - 0.0293i, -0.0655 - 0.0124i, -0.0664 + 0.0055i, -0.0626 + 0.0229i, -0.0543 + 0.0387i]

$w3$ =[0.1000, 0.0964 - 0.0265i, 0.0859 - 0.0512i, 0.0692 - 0.0721i, 0.0476 - 0.0879i, 0.0226 - 0.0974i, -0.0041 - 0.0999i, -0.0305 - 0.0952i, -0.0547 - 0.0837i, -0.0749 - 0.0662i]

2) Effect of spacing between elements on array factor

The effect of array spacing for $\lambda / 2$, $\lambda / 4$ and $\lambda / 8$ is shown in Fig. 6 for $Ne = 10$. Since the spacing between the elements is critical, due to sidelobes problems, which causes spurious echoes and diffraction secondaries, which are repetitions of the main beam within the range of real angles.

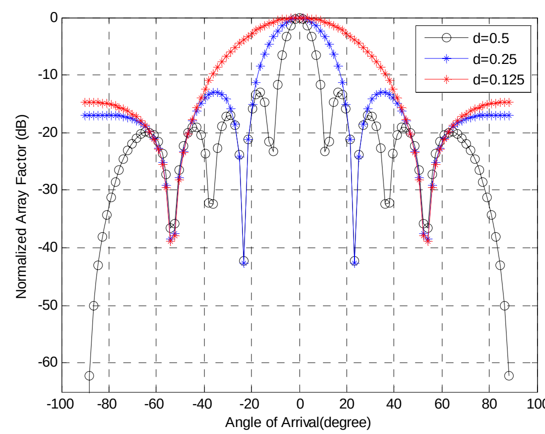


Fig.6. Normalized array factor plot for MVDR algorithm for $Ne = 10$ **with interferer – 50 degrees**

From Fig. 6, it is observed that increasing element spacing produces narrower beams, but this happens at the cost of increasing number of sidelobes. It is also clear, that spacing between elements equal to $\lambda / 2$, gives optimum result for narrower beam.

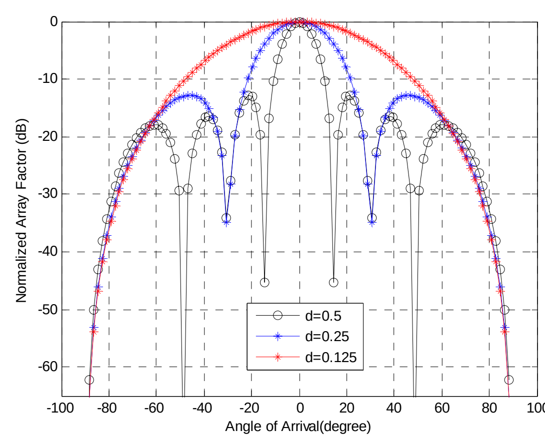


Fig.7. Normalized Array factor plot for MVDR algorithm for $Ne = 8$ **with interferer – 30 degrees**

When number of elements is reduced to 8, then effect of array spacing is shown at Fig. 7. Again, narrower beam width is achieved at $d = \lambda / 2$.

III. KAPA ALGORITHM

A. *Theory*

For the first time, KAPA algorithm is presented in [1],

.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

for noise cancellation and providing a unifying model for several neural networks techniques. It is the combination of famed kernel trick and affine projection (APA) algorithm [10]. In our case, this algorithm is employed for beamforming which is novel in this application [11]. In KAPA algorithm, the input signal $u(n)$ is transformed into a high dimensional feature space via a positive definite kernel such that the inner product operation in the feature space can be computed efficiently through the kernel evaluation. KAPA is categorized as nonblind algorithm which uses a desired/training signal to update its complex weights vector. This training signal is sent by the transmitter to the receiver during the training period.

The weight $w(n)$ update equation for the KAPA algorithm is defined as:

$$w(n) = w(n-1) + \eta\varphi(n)\varepsilon(n)$$

$$= \sum_{n=1}^{k-1} a_n(k-1)\varphi(n) + \sum_{n=1}^{K} \eta\varepsilon_n(n)\varphi(n-1+K) \quad (13)$$

where φ is an eigen functions, ε is a positive regularization factor and η is the step size.

During the iteration, the weight vector in the feature space assumes the following expansion as given by

$$w(n) == \sum_{n=1}^{k} a_n(k)\varphi(n)\forall_n > 0 \quad (14)$$

That is, the weight at time n is a linear combination of the previous transformed input.

The error signal is computed by

$$\varepsilon(n) = d(n) - \phi(n)w(n-1) \quad (15)$$

where $d(n)$ is the desired signal, used for training sequence of known symbols (also called a pilot signal), is required to train the adaptive weights. Enough training sequence of known symbols must be available to ensure convergence [4] [5] [9].

B. *Simulation Results*

1) *Effect of number of elements on array factor*

Uniform linear array is taken for simulation purpose. AOA for desired user is set at 0 & 20 degrees and for interferer at – 50 & – 20 degrees as shown in Fig 8 and 9, respectively. The space $(\lambda/8)$ is maintained between elements. The narrow beam with side lobes is observed for longer array.

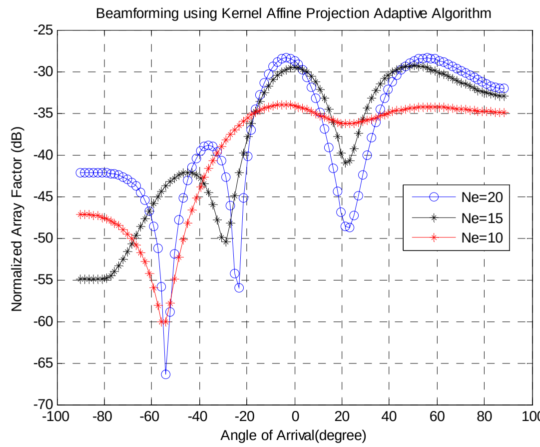


Fig.8. Normalized array factor plot for KAPA algorithm with AOA for desired user is 0 degree and **- 50 degrees for interferer** with constant space of $(\lambda/8)$ between elements

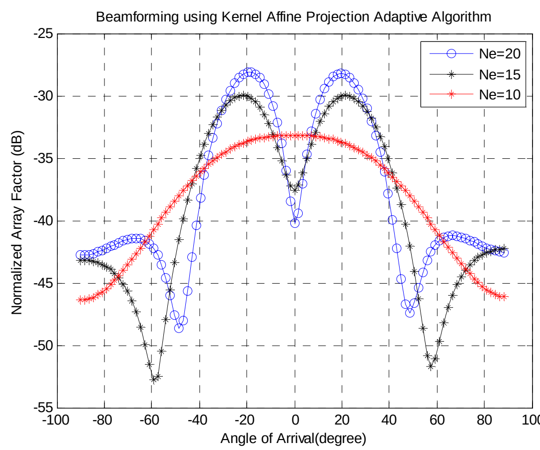


Fig.9. Normalized array factor plot for KAPA algorithm with AOA for desired user is 20 degrees and **- 20 degrees for interferer** with constant space of $(\lambda/8)$ between elements

Now if number of elements is changed then broad beam is obtained with reduced sidelobes as shown in Fig. 10, for desired user at 20 degrees and for interferer is at - 40 degrees.

.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

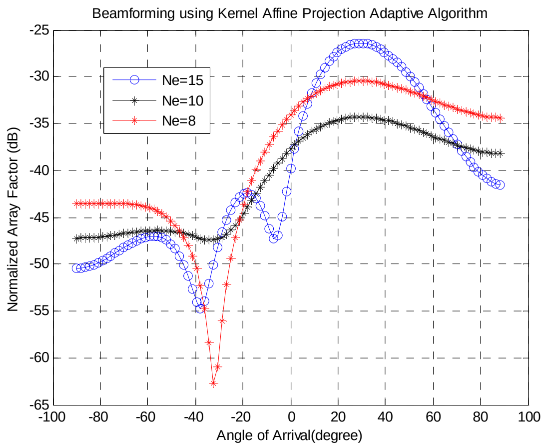


Fig.10. Normalized array factor plot for KAPA algorithm with AOA for desired user is 20 degrees and **- 40 degrees for interferer** with constant space of $(\lambda/8)$ between elements

The weight vectors obtained during convergence for $Ne = 20,15$ and 10 are $w1, w2$ and $w3$, respectively as shown in Fig. 11. Numerically, these weight vectors are represented as:

$w1$=[0.0030 + 0.0016i, 0.0036 + 0.0002i, 0.0035 - 0.0011i, 0.0030 - 0.0026i, 0.0016 - 0.0034i, 0.0006 - 0.0039i, -0.0008 - 0.0036i, -0.0024 - 0.0030i, -0.0030 - 0.0017i, -0.0030 - 0.0004i, -0.0028 + 0.0006i, -0.0023 + 0.0015i, -0.0012 + 0.0020i, -0.0003 + 0.0019i, 0.0003 + 0.0015i]

$w2$=[0.0033 - 0.0020i, 0.0018 - 0.0025i, 0.0009 - 0.0027i, -0.0003 - 0.0025i, -0.0007 - 0.0020i, -0.0014 - 0.0012i, -0.0015 - 0.0003i, -0.0010 + 0.0001i, -0.0005 + 0.0001i, 0.0001 - 0.0000i]

$w3$=[0.0031 + 0.0017i, 0.0037 + 0.0004i, 0.0036 - 0.0010i, 0.0030 - 0.0026i, 0.0019 - 0.0033i, 0.0003 - 0.0037i, -0.0009 - 0.0040i, -0.0020 - 0.0028i]

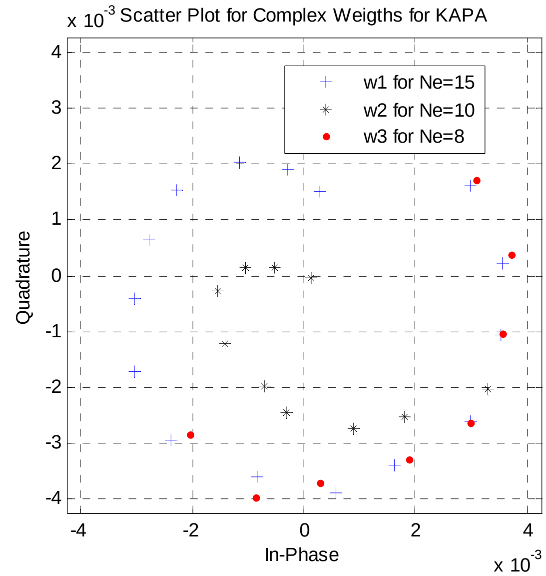


Fig.11. Scatter plot for **complex weights** for $Ne = 20,15$ and 10 with constant space of $(\lambda/8)$ between elements

2) Effect of spacing between elements on array factor

When number of elements is kept constant for different array spacing i.e. $d = \lambda/2$, $d = \lambda/4$ and $d = \lambda/8$, then its effect is shown in Fig. 12 and 13 for $Ne = 10$ and $Ne = 8$, respectively. The sharp beam is obtained for $Ne = 10$ for $d = \lambda/2$ as compared to $Ne = 8$. AOA for desired user is set at 0 and - 60 degrees for interferer in Fig.12 but deep null is observed at 50 degree instead of - 60 degree.

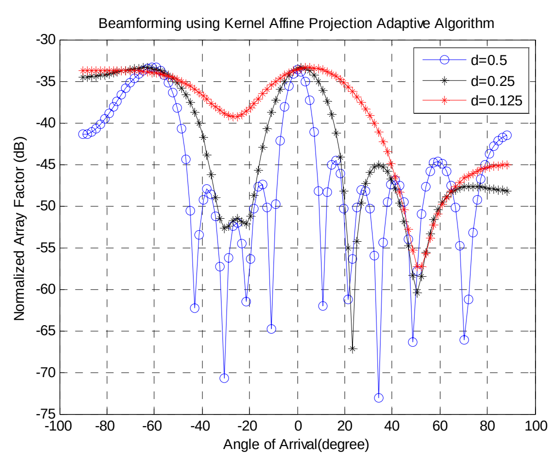


Fig.12. Normalized Array factor plot for KAPA algorithm for $Ne = 10$ **with interferer – 60 degrees**

.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Similarly AOA for desired user is set at - 20 and - 70 degrees for interferer in Fig.13 but deep null is observed at 40 degree instead of - 70 degree.

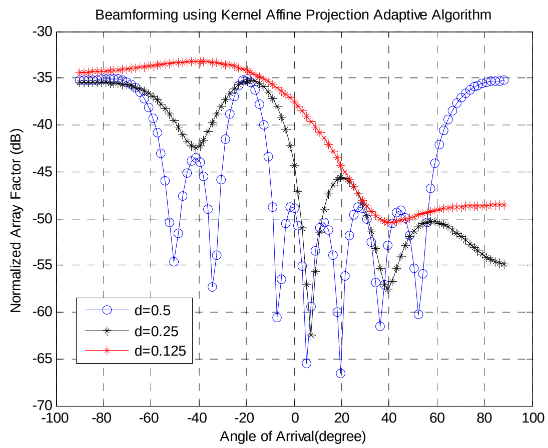


Fig.13. Normalized Array factor plot for KAPA algorithm for $Ne = 8$ **with interferer – 70 degrees**

IV. COMPARISON ON THE BASIS OF AOA

MVDR and KAPA algorithms can also be compared on the basis of AOA as shown in Fig. 14 and 15, respectively. Both these algorithms have shown best response for beamforming keeping $(\lambda/8)$ spacing between elements.

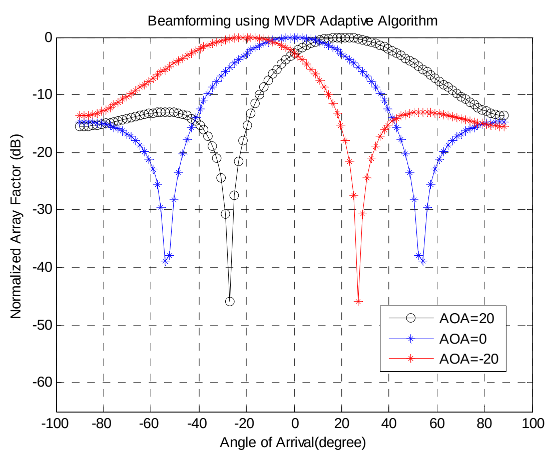


Fig.14. Normalized Array factor plot for MVDR algorithm for $Ne = 10$

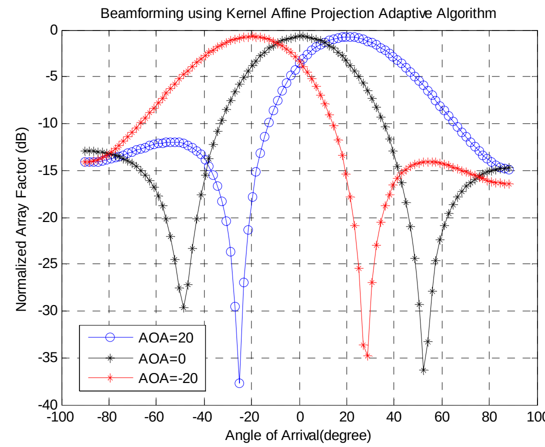


Fig.15. Normalized Array factor plot for KAPA algorithm for $Ne = 10$

V. CONCLUSIONS

The performance analysis of blind algorithm that is MVDR and nonblind algorithm i.e. KAPA is carried out in this paper. These algorithms are used in smart/adaptive antenna array system in coded form to generate beam in the look direction and null towards interferer, thus enhancing performance of mobile communication systems in terms of channel capacity, tailoring beam shape and steering beams to track many mobiles electronically It is confirmed from the simulation results that narrow beam of smart antenna can be steered towards the desired direction by steering beam angle Φ_0, keeping elements spacing d, number of elements Ne and altering weights $w(n)$ adaptively for both algorithms. However, MVDR algorithm has shown better response towards desired direction and has good capability to place null towards interferer as compared to KAPA. The convergence speed of MVDR algorithm is better as it does not rely on eigen values whereas KAPA depends on eigen functions, therefore its speed of convergence is slow as compared to MVDR. It is also ascertained from the simulation results that MVDR algorithm has shown better performance in beam formation taking different number of elements and for different spacing maintained between elements. However, KAPA algorithm has exercised reasonable performance inculcation of beampattern for same number of iteration and for same parameters being used for MVDR. It is worth noting that MVDR is simple in computation as it doesn't require training signal for convergence as compared to KAPA. Therefore, maximum bandwidth is utilizing to exchange information between transmitters and receivers, thus enhancing capacity. Keeping these advantages in mind, MVDR is found a better option to implement at base

.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

station of mobile communication systems using CDMA environment to reduce interference, enhance capacity and service quality.

REFERENCES

[1] Weifeng Liu and Jose C. Principe, "Kernel affine projection algorithms," EURASIP Journal on Advances in Signal Processing, VOL. 2008, Article ID 784292, 12 pages, 21 February 2008.
[2] LAL. C. GODARA, Senior Member, IEEE, "Applications of antenna arrays to mobile communications, Part I; performance improvement, feasibility, and system considerations," Proceeding of the IEEE, VOL. 85, NO. 7, pp. 1031-1060, July 1997.
[3] LAL. C. GODARA, Senior Member, IEEE, "Applications of antenna arrays to mobile communications, Part II; beam-forming and directional of arrival considerations," Proceeding of the IEEE, VOL. 85, NO. 8, pp1195-1245, August 1997.
[4] Simon Haykin, Adaptive Filter Theory, Fourth edition (Pearson Eduation, Inc., 2002).
[5] B. Widrow and S.D. Stearns, Adaptive Signal Processing (Pearson Eduation, Inc., 1985).
[6] Kalavai J. Raghunath, Student Member, IEEE, V. Umapathi Reedy, Senior Member, IEEE, "Finite data performance analysis of MVDR beamformer with and without spatial smoothing," IEEE Transactions on Signal Processing, VOL. 40, NO. 11, pp2126-2136, November 1992.
[7] F. E. Fakoukakis, S. G. Diamantis, A. P. Orfanides and G. A. kyriacou, "Development of an adaptive and a switched beam smart antenna system for wireless communications," progress in electromagnetics research symposium 2005, Hangzhou, China, pp. 1-5, August 22-26, 2005.
[8] Rameshwar Kawitkar, "Issues in deploying smart antennas in mobile radio networks," Proceedings of World Academy of Science, Engineering and Technology Volume 31 July 2008, pp. 361-366, ISSN 1307-6884.
[9] Hun Choi and Hyeon-Deok Bae, "Subband affine projection algorithm for acoustic echo cancellation system," EURASIP Journal on Advances in Signal Processing, VOL. 2007, Article ID 75621, 12 pages, 18 May 2006.
[10] M Yasin, Pervez Akhtar and M Junaid Khan, "Affine Projection Adaptive Filter a Better Noise Canceller," IST Journal CSTA, in press.
[11] M Yasin, Pervez Akhtar and M Junaid Khan, "Tracking Performance of RLS and KAPA Algorithms for a Smart Antenna System," unpublished.
[12] M Yasin, Pervez Akhtar and Valiuddin, "Performance Analysis of LMS and NLMS Algorithms for a Smart Antenna System," Journal IJCA, in press.
[13] M Yasin, Pervez Akhtar and M Junaid Khan, "CMA an Optimum Beamformer for a Smart Antenna System," Journal IJCA, in press.

Muhammad Yasin is enrolled for PhD in the field of electrical engineering majoring in telecommunication in Pakistan Navy Engineering College, National University of Science and Technology, Karachi (NUST), Pakistan. He is working in Pakistan Navy as naval officer in the capacity of communication engineer since 1996. His research interests include signal processing, adaptive filtering, implementation of communication networking and its performance evaluation. He has received a B.Sc. degree in electrical engineering from NWFP University of Engineering and Technology, Peshawar (1994) and M.Sc. degree in electrical engineering from NED, University of Engineering and Technology, Karachi (2006). He has also done a Master degree in Economics (2002) from University of Karachi. In the past, he is involved in implementation of ISO 9000 on indigenous project of AGOSTA 90B Class Submarines along with French engineers. Currently, he is working on indigenous project of Acoustic System Trainer, being used for imparting Sonar related training.

.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

(IJCSIS) International Journal of Computer Science and Information Security,
Vol. 8, No. 4, July 2010

Specifying And Validating Quality Characteristics For Academic Web-sites – Indian Origin

Ritu Shrivastava
Department of Computer Science and Engineering
Sagar Institute of Research Technology & Science
Bhopal 462007, India

J. L. Rana
Retired Professor, Department of Computer Science and Engineering
Maulana Azad National Institute of Technology
Bhopal 462002, India

M Kumar
Prof. & Dean, Department of Computer Science and Engineering
Sagar Institute of Research & Technology
Bhopal 462007, India

Abstract— **Every stakeholder of Academic Web-sites is mainly concerned with external quality, viz., usability, functionality, and reliability. Signore and Olsina have given hierarchical quality characteristics for measuring quality of Web-sites, especially for e-commerce and museum domains. In this paper, the authors have proposed a hierarchical model of attributes, sub-attributes, and metrics for measuring external quality of academic Web-sites – Indian origin. The theoretical validation of model has been carried out using distance measure construction method. The empirical validation is in progress and will be reported soon.**

Keywords-component; Web-site Quality, Academic domain, Hierarchical model, Attributes, Metrics

I. INTRODUCTION

World Wide Web (WWW) has been the fasted adopted technology. Every day many new Web-sites are uploaded on Web. Often quality of Web-sites is unsatisfactory and basic Web principles like inter-portability and accessibility are ignored [1,2]. The main reason for lack of quality is unavailability of trained staff in Web technologies/engineering and orientation of Web towards a more complex XML based architecture [1,2,3].

Web-sites can be categorized as informative or cultural, e-commerce, e-government, museums, tourism, and academic intensive. It is obvious that domains differ significantly, and hence a common yardstick can not be applied to measure quality of all Web-sites. Loranca et. al. [4] and Olsina et. al. [5] have identified attributes, sub-attributes, and metrics for e-commerce based Web-sites. Olsina et. al. [6] have also specified metrics for Web-sites of museums. Tripathi and Kumar [7] have specified quality characteristics for e-commerce based Web-sites of Indian origin from external point of view.

The aim of this research is to identify attributes, sub-attributes, and metrics for measuring quality of Academic Institute Web-sites (Indian Origin) from point of view of usability, and to theoretically validate the proposed model.

II. LITRATURE SURVEY

The quality of software being developed has always been prime concern of software engineers. Some widely used software quality models were proposed by Boehm et. al. [8], and McCall and Covano [9]. Complexity is probably the most important attribute of software because it influences a number of other attributes such as maintainability, understandability, modifiability, and testability.

International bodies such as ISO and CEN(European) are trying to integrate different approaches to the definition of quality, starting from the awareness that the quality as an attribute which changes developer's perspective and action context [10]. The ISO/IEC 9126 model [10] defines three views of quality: user's view, developer's view, and manager's view. Users are interested in the quality in use (external quality attributes), while developers are interested in internal quality attributes such as maintainability, portability, etc.. This model is hierarchical and contains six major quality attributes each very broad in nature. They are subdivided into 27 sub-attributes that contribute to external quality and 21 sub-attributes that contribute to internal quality. The users are interested in external quality, viz., usability, functionality, reliability, and efficiency of Web-sites. These attributes and sub-attributes in ISO 9126 are of very general in nature and can be applied to Web-sites as well.

Olsina et. al.[5,6] have proposed hierarchical model of attributes, sub-attributes and metric for Web-sites of museum and e-commerce domains. They have also developed a technique called QEM to measure quality of these sites [5]. Tripathi and Kumar [7] have identified attributes, sub-attributes and metrics for Indian origin e-commerce Web-sites. They have validated the proposed quality characteristic model both theoretically and empirically [11]. In this research we are

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

proposing a hierarchical model of attributes and sub-attributes to measure quality of academic institute Web-sites of Indian origin. The model is also theoretically validated.

III. PROPOSED QUALITY CHARACTERISTICS MODEL

In fact, software artifacts are generally produced to satisfy specific user's need, and Web-sites are no exception. In designing Web-sites care should be taken that a user entering for the first time at a given home page should be able to find a piece of information quickly. For this, there are attributes like site map, an index, or a table of contents that help in getting quick global site understanding that facilitates browsing. Alternatively, a global searching function on the home page could help retrieving required piece of information and avoid browsing. The site understandability increases if both the functions are included. The main attributes that enhance the Web-site external quality are usability, functionality, and reliability. A quality attribute can be decomposed into multiple levels of sub-attributes and finally a sub-attribute can be refined in a set of measurable attributes or metrics. The proposed hierarchical model of metrics to measure external quality of academic Web-sites is given in Fig. 1.

It is necessary that any new model of attributes, sub-attributes and metrics is properly validated before it is put to use by professionals and academia. The process of validation is described in the next section.

IV. THEORETICAL VALIDATION OF PROPOSED HIERARCHICAL MODEL OF METRICS

Recent software engineering literature has shown a concern for the quality of methods to validate software product metrics (e.g., see [12][13][14]). This concern is due to fact that: (i) common practices for the validation of software engineering metrics are not acceptable on scientific grounds, and (ii) valid measures are essential for effective software project management and sound empirical research. According to Kitchenham et al. [13] "unless the software measurement community can agree on a valid, consistent, and comprehensive theory of measurement validation, we have no scientific basis for the discipline of software measurement, a situation potentially disastrous for both practice and research." Therefore, to have confidence in the utility of the many metrics those are proposed from research labs, it is crucial that they are validated.

The validation of software product metrics means

1 Usability
- 1.1. Global Site understandability
  - 1.1.1 *Site Map(location map)*
  - 1.1.2 *Table of Content*
  - 1.1.3 *Alphabetical Index*
  - 1.1.4 *Campus Image Map*
  - 1.1.5 *Guided Tour*
- 1.2. On-line Feedback and Help Features
  - 1.2.1 *Student Oriented Help*
  - 1.2.2 *Search Help*
  - 1.2.3 *Web-site last Update Indicator*
  - 1.2.4 *E-mail Directory*
  - 1.2.5 *Phone Directory*
  - 1.2.6 *FAQ*
  - 1.2.7 *On-line Feedback in form of Questionnaire*
- 1.3. Interface and Aesthetic Features
  - 1.3.1 *Link Color Style Uniformity*
  - 1.3.2 *Global Style Uniformity*
  - 1.3.3 *What is New Feature*
  - 1.3.4 *Grouping of Main Control Objects*

2 Functionality
- 2.1. Search Mechanism
  - 2.1.1 *People Search*
  - 2.1.2 *Course Search*
  - 2.1.3 *Academic Department Search*
  - 2.1.4 *Global Search*
- 2.2. Navigation and Browsing
  - 2.2.1 *Path Indicator*
  - 2.2.2 *Current Position Indicator*
  - 2.2.3 *Average Links Per Page*
  - 2.2.4 *Vertical Scrolling*
  - 2.2.5 *Horizontal Scrolling*
- 2.3. Student-Oriented Features
  - 2.3.1 Academic Infrastructure Information
    - 2.3.1.1 *Library Information*
    - 2.3.1.2 *Laboratory Information*
    - 2.3.1.3 *Research Facility Information*
    - 2.3.1.4 *Central Computing Facility Information*
  - 2.3.2 Student Service Information
    - 2.3.2.1 *Hostel Facility Information*
    - 2.3.2.2 *Sport Facilities*
    - 2.3.2.3 *Canteen Facility Information*
    - 2.3.2.4 *Scholarship Information*
    - 2.3.2.5 *Doctor/Medical Facility Information*
  - 2.3.3 Academic Information
    - 2.3.3.1 *Courses Offered Information*
    - 2.3.3.2 *Academic Unit (Department) Information*
    - 2.3.3.3 *Academic Unit t Site Map*
    - 2.3.3.4 *Syllabus Information*
    - 2.3.3.5 *Syllabus Search*
  - 2.3.4 Enrollment Information
    - 2.3.4.1 *Notification uploaded*
    - 2.3.4.2 *Form Fill/Download*
  - 2.3.5 Online Services
    - 2.3.5.1 *Grade/ Result Information*
    - 2.3.5.2 *Fee dues/Deposit Information*
    - 2.3.5.3 *News Group Services*

3 Reliability
- 3.1. Link and Other Errors
  - 3.1.1 *Dangling Links*
  - 3.1.2 *Invalid Links*
  - 3.1.3 *Unimplemented Links*
  - 3.1.4 *Browser Difference Error*
  - 3.1.5 *Unexpected Under Construction Pages*

4 Efficiency
- 4.1 Performance
  - 4.1.1 *Matching of Link Title and Page Information*
  - 4.1.2 *Support for Text only Version*
  - 4.1.3 *Global Readability*
  - 4.1.4 *Multilingual Support*

Fig. 1 Quality Characteristics For Academic Institute Web-sites

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

convincingly demonstrating that :

1. The product metrics measures what it purports to measure. For example, that a coupling metrics really measures coupling.

2. The product metric is associated with some important external metric (such as measures of maintainability or reliability).

3. The product metric is an improvement over existing product metrics. An improvement can mean, for example, that it is easier to collect the metric or that it is a better predictor of faults.

According to Fenton [15], there are two types of validation that are recognized: internal and external. Internal validation is a theoretical exercise that ensures that the metric is a proper numerical characterization of the property it claims to measure. Demonstrating that a metric measures what it purports to measure is a form of theoretical validation. External validation involves empirically demonstrating points (2) and (3) above. Internal and external validations are also commonly referred to as theoretical and empirical validation respectively [13]. Both types of validation are necessary. The approaches used in two validations are shown in Figure 2.

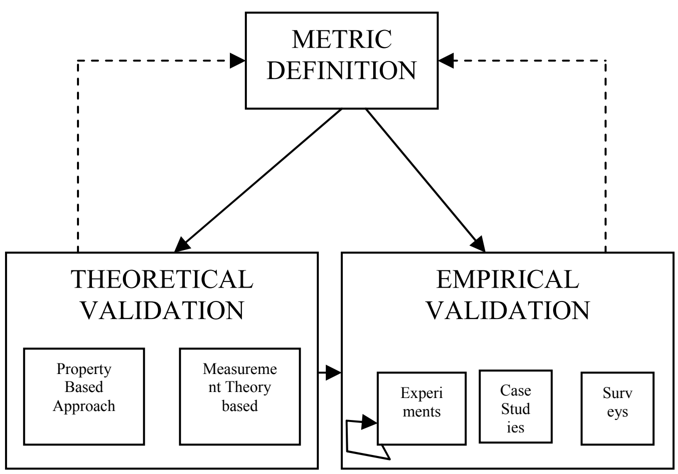


Fig. 2 Approaches to Software Metrics Validation

The main goal of theoretical validation is to assess whether a metric actually measures what it purports to measure [15]. In the context of an empirical study, the theoretical validation of metrics establishes their construct validity, i.e. it 'proves' that they are valid measures for the constructs that are used as variables in the study. There is not yet a standard, accepted way of theoretically validating software metric. Work on theoretical validation has followed two paths (see Fig 2), viz.

- Measurment-theory based approach such as those proposed by Whitmire[16], Zuse[17], and Poels and Dedene [18]
- Property-based approach (also called axiomatic approaches), such as proposed by Weyuker and Braind et al.[19]

For the theoretical validation DISTANCE framework proposed by Poels and Dedene[18], is a conceptual framework for software metric validation grounded in measurement theory. This is briefly described below :

A. The DISTANCE Measure Construction Procedure

The measure construction procedure prescribes five activities. The procedure is triggered by a request to construct a measure for a property that characterizes the element of some set of objects. The activities of the DISTANCE procedure are given here. For notational convenience, let P be a set of objects that are characterized by some property pty for which a measure needs to be constructed.

1) Finding a measurement abstraction: The object of interest must be modeled in such a way that the property for which a measure is needed is emphasized. A suitable representation, called measurement abstraction hereafter, should allow to what extent an object is characterized by the property to be observed. By comparing measurement abstraction we should be able to tell whether an object is more, equally or less characterized by the property than other object.

2) Defining distance between measurement abstraction: This activity is based on a generic definition of distance that hold for elements in a set. To define distance between elements in a set, the concept of 'elementary transformation function' is used.

3) Quantifying distance betweenmeasurement abstraction: This activity requires the definition of a distance measure for the element of M. Basically this means that the distance defined in the previous activity are now quantified by representing i.e. measuring them as the number of elementary transformation by representing i.e. measuring them as the number of elementary transformations in the shortest sequence of elementary transformation between elements. Formally, the activity results in the definition of a metric $MxM \rightarrow R$ that can be used to map the distance between a pair of elements in M to a real number.

4) Finding a reference abstraction: This activity require a kind of thought experiment. We need to determine what the measurement abstraction for the object in P would look like if they were characterized by the theoretical lowest amount pty. If such a hypothetical measurement abstraction can be found, then this object is called the reference abstraction for P with respect to pty.

5) Defining a measure for the property: The final activity consists of defining a measure for pty. Since properties are formally defined as distances, and these distances are quantified with a metric function, the formal outcome of this activity is the definition of a function $\mu:P \rightarrow R$ such that $p \in P$: $\mu(p) = \delta(abs(p), ref(p))$.

B. Metric Validation

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

The proposed hierarchical model of metrics given in Fig 2 is validated using Distance methodology. We have used the five activities of DISTANCE measure procedure for metrics of the model and important metrics are summarized in Table 1

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

TABLE I. DISTANCE BASED VALIDATION CRITERIA FOR METRICS

Quality	Metrics	Validation				
		Measurement Abstraction	Defining distance between two extreme abstraction	Quantifying distance e in extremes	Hypothetical reference abstraction	Determining a measure ptv
Usability	1.1.1	Nominal	Yes/No	EQ = {1,0}	EQ=0, if no map	EQ=1, if map available
	1.1.2	Nominal	Yes/No	EQ = {1,0}	EQ=0, if no table	EQ=1, if table available
	1.1.3	Nominal	Yes/No	EQ = {1,0}	EQ=0, if no alphabetic index	EQ=1, if alphabetic index available
	1.1.4	Nominal	Yes/No	EQ = {1,0}	EQ=0, if no image map	EQ=1, if image map available
	1.1.5	Nominal	Yes/No	EQ = {1,0}	EQ=0, if no guided tour	EQ=1, if guided tour available
	1.2.1	Nominal	Yes/No	EQ = {1,0}	EQ=0, if no student oriented help	EQ=1, if student oriented help available
	1.2.2	Nominal	Yes/No	EQ = {1,0}	EQ=0, if no search help	EQ=1, if search help available
	1.2.3	Nominal	Yes/No	EQ = {1,0}	EQ=0, if no update indicator	EQ=1, if update indicator available
	1.2.4	Ordinal	Complete/Partial/No	EQ = {1,0.5,0}	EQ=0, if no email directory	EQ=1, if complete email directory available
	1.2.5	Ordinal	Complete/Partial/No	EQ = {1,0.5,0}	EQ=0, if no phone directory	EQ=1, if complete phone directory available
	1.2.6	Ordinal	Exhaustive/Partial/No	EQ = {1,0.5,0}	EQ=0, if no FAQ	EQ=1, if exhaustive FAQs available
	1.2.7	Nominal	Yes/No	EQ = {1,0}	EQ = 0, if no feedback	EQ=1, if feedback available available
	1.3.1	Ordinal	Uniform/Partial/No	EQ = {1,0.5,0}	EQ=0, if no link color style	EQ=1, if uniform link color style available
	1.3.2	Ordinal	Uniform/Partial/No	EQ = {1,0.5,0}	EQ=0, if no global style uniformity	EQ=1, if global style uniformity available
	1.3.3	Nominal	Yes/No	EQ = {1,0}	EQ=0, if no new feature	EQ=1, if new feature available
	1.3.4	Ordinal	Complete/Partial/No	EQ = {1,0.5,0}	EQ=0, if no grouping of objects	EQ=1, if complete grouping of objects
Functionality	2.1.1	Nominal	Yes/No	EQ = {1,0}	EQ=0, if no people search	EQ=1, if people search available
	2.1.2	Nominal	Yes/No	EQ = {1,0}	EQ=0, if no course search	EQ=1, if course search available
	2.1.3	Nominal	Yes/No	EQ = {1,0}	EQ=0, if no department search	EQ=1, if department search available
	2.1.4	Nominal	Yes/No	EQ = {1,0}	EQ=0, if no global search	EQ=1, if global search available
Functionality	2.2.1	Nominal	Yes/No	EQ = {1,0}	EQ = 0, if no path indicator	EQ=1, if path indicator available
	2.2.2	Nominal	Yes/No	EQ = {1,0}	EQ = 0, if no current position	EQ=1, if current position available
	2.2.3	Ordinal	Good/Average/Bad	EQ = {1,0.5,0}	EQ = 0, if no average link per page	EQ=1, if average link per page 6 or more
	2.2.4	Nominal	Yes/No	EQ = {1,0}	EQ = 0, if no vertical scrolling	EQ=1, if vertical scrolling available
	2.2.5	Nominal	Yes/No	EQ = {1,0}	EQ = 0, if no horizontal scrolling	EQ=1, if horizontal scrolling available
	2.3.1.1	Nominal	Yes/No	EQ = {1,0}	EQ = 0, if no library info	EQ=1, if library Info available
	2.3.1.2	Nominal	Yes/No	EQ = {1,0}	EQ = 0, if no laboratory info	EQ=1, if laboratory Info available
	2.3.1.3	Nominal	Yes/No	EQ = {1,0}	EQ = 0, if no research facility	EQ=1, if research facility available
	2.3.1.4	Nominal	Yes/No	EQ = {1,0}	EQ = 0, if no central computing facility	EQ=1, if central computing facility available
	2.3.2.1	Nominal	Yes/No	EQ = {1,0}	EQ = 0, if no hostel facility info	EQ=1, if hostel facility info available
	2.3.2.2	Nominal	Yes/No	EQ = {1,0}	EQ = 0, if no sports	EQ=1, if sports facility

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Quality	Metrics	Validation				
		Measurement Abstraction	Defining distance between two extreme abstraction	Quantifying distance e in extremes	Hypothetical reference abstraction	Determining a measure ptv
					facility info	info available
	2.3.2.3	Nominal	Yes/No	EQ = {1,0}	EQ = 0, if no canteen facility info	EQ=1, if canteen facility info available
	2.3.2.4	Nominal	Yes/No	EQ = {1,0}	EQ = 0, if no scholarship info	EQ=1, if scholarship info available
	2.3.2.5	Nominal	Yes/No	EQ = {1,0}	EQ = 0, if no medical facility info	EQ=1, if medical facility info available
	2.3.3.1	Nominal	Yes/No	EQ = {1,0}	EQ = 0, if no courses offered info	EQ=1, if courses offered info available
	2.3.3.2	Nominal	Yes/No	EQ = {1,0}	EQ = 0, if no department info	EQ=1, if department info available
	2.3.3.3	Nominal	Yes/No	EQ = {1,0}	EQ = 0, if no Dept. site map	EQ=1, if Dept. site map available
	2.3.3.4	Nominal	Yes/No	EQ = {1,0}	EQ = 0, if no syllabus info	EQ=1, if syllabus info available
	2.3.3.5	Nominal	Yes/No	EQ = {1,0}	EQ = 0, if no syllabus search	EQ=1, if syllabus search available
	2.3.4.1	Nominal	Yes/No	EQ = {1,0}	EQ = 0, if no notifications	EQ=1, if notifications available
	2.3.4.2	Nominal	Yes/No	EQ = {1,0}	EQ = 0, if no form download	EQ=1, if form download available
Functionality	2.3.5.1	Ordinal	Complete/Partial/No	EQ = {1,0.5,0}	EQ = 0, if no result info	EQ=1, if all sem result info available
	2.3.5.2	Ordinal	Complete/Partial/No	EQ = {1,0.5,0}	EQ = 0, if no fee dues info	EQ=1, if full fee dues info available
	2.3.5.3	Nominal	Yes/No	EQ = {1,0}	EQ = 0, if no news group	EQ=1, if news group available
Reliability	3.1.1	Nominal	Yes/No	EQ = {1,0}	EQ = 0, if no dangling link	EQ=1, if dangling link available
	3.1.2	Ordinal	High /Medium/ Low	EQ={1, if bet 0-2; 0.5, if bet 3-5; 0, if 6 or more}	EQ = 0, if invalid links 6 or more	EQ=1, if invalid links bet 0-2
	3.1.3	Ordinal	High /Medium/ Low	EQ={1, if bet 0-2; 0.5, if bet 3-5; 0, if 6 or more}	EQ = 0 if unimplemented link 6 or more	EQ=1, if unimplemented links bet 0-2
	3.1.4	Nominal	Yes/No	EQ = {1,0}	EQ = 0, if no browser difference	EQ=1, if browser difference available
	3.1.5	Ordinal	High /Medium/ Low	EQ={between 0-2, 3-5, 6 or more}	EQ = 0 if unconstructed pages 6 or more	EQ=1, if unconstructed bet 0-2
Efficiency	4.1.1	Ordinal	High /Medium/ Low	EQ={between 0-2, 3-5, 6 or more}	EQ = 0 if unmatching link 6 or more	EQ=1, if unmatching link bet 0-2
	4.1.2	Nominal	Yes/No	EQ = {1,0}	EQ = 0, if no support	EQ=1, if support for text only version
	4.1.3	Nominal	Yes/No	EQ = {1,0}	EQ = 0 if no global reliability	EQ=6, if global reliability good
	4.1.4	Nominal	Yes/No	EQ = {1,0}	EQ = 0, if no multilingual support	EQ=1, if multilingual support available

V. CONCLUSION

We have proposed a hierarchical model of attributes, sub-attributes, and metrics for measuring quality of Indian origin academic Web-sites from the point of view of usability, which is of major concern to users (stake holders). The proposed metrics are theoretically validated using distance measure construction procedure and results are shown in the Table 1. The empirical validation is in progress and will be reported soon.

VI. ACKNOWLEDGMENT

Authors sincerely thank Dr R. K. Pandey, Director, Institute of Technology, Barkatullah University, Bhopal for his constant support and expert guidance during preparation of this paper.

REFRENCES

[1] O. Signore , “ Towards a quality model for Web-sites” , CMG Poland Annual Conference, Warsaw, 9-10 May, 2005, http://www.w3c.it/papers/cmg2005Poland-quality.pdf.

[2] J. Offutt , " Quality attributes of Web software applications" , IEEE Software, March/April, pp25-32, 2002.

[3] O. Signore, et. al. , "Web accessibility principles" , I*nternational Context and Italian Regulations", EuroCMG, Vienna, 19-21 Sept. 2004, http://www.w3c.it/paperseurocmg2004.pdf.*

[4] M. B. Loranca, J. E.Espinosa, et. al. , "Study for classification of quality attributees in Argentinean E-commmerce sites" , Proc. 16th IEEE Itern. Conf. on Electronics Communication & Computers 2006.

[5] L. Olsina and G. Rossi, "Measuring Web application quality with WebQEM" , IEEE Multimedia, pp 20-29. Oct-Dec 2002.

[6] L. Olsina , "Website quality evaluation method : A case study of Museums", 2nd workshop on Software Engineering over Iternet, ICSE 1999.

[7] P. Tripathi, M. Kumar ," Some observations on quality models for Web-applications" , Proc. of Intern Conf on Web Engineering and Applications, Bhubaneshwar, Orissa, India, 23-24 Dec 2006 (Proc Published by Macmillan 2006).

[8] B. Boehm, J. Brown, M. Lipow, "Quantitative evaluation of software quality process" , Intern. Conference on Software Engineering, IEEE Computer Society Press, pp 592-605, 1976.

[9] J. Covano, J. McCall , "A framework for measurement of software quality " , Proc. ACM Software Quality Assurance Workshop, pp133-139, 1978.

[10] ISO/IEC 9126-1 : Software Engineering – Product Quality Part 1 : Quality Model(2000) : http://www.usabilitynet .org/tools/international.html#9126-1.

[11] P. Tripathi , M. Kumar and N. Shrivastava, " Ranking of Indian E-commerce Web-applications by measuring quality factors " , Proc of 9th ACIS Itern Conf on Software Engineering, AI, Networking and Parallel/Distributed Computing, Hilton Phulket, Thailand, Aug 6-8, 2008. Proc Published by IEEE Comp. Soc.

[12] V. Fenton, B. Kitchenham, "Validating software measures", Journal of Software Testing, Verification and Reliability, vol. 1, no. 2, pp. 27-42, 1990.

[13] B. Kitchenham, S-L Pfleeger, and N. Fenton, "Towards a framework for software measurement validation" , IEEE Transactions on Software Engineering, vol 21, no. 12, pp. 929-944, 1995.

[14] N. Schneidewind, "Methodology for validating software metrics," IEEE Transactions on Software Engineering, vol. 18, no. 5, pp. 410-422, 1992.

[15] N.Fenton, "Software metrics: theory, tools and validation, "SoftwareEngineering Journal, pp. 65-78, January, 1990.

[16] J. Whitmire, "Correctly assigning the "ilities" requires more than marketing hype", IT Professional , vol 2, no 6, pp 65-67, 2000.

[17] H. Zuse, A framework of software measurement , Walter de Gruyter, Berlin, 1998.

[18] G. Poels and G. Dedene, "Distance-based software measurement: necessary & sufficient properties for software measures ", Information and Software Technology, vol 42, no 1, pp 35-46, 2000

[19] L. Briand, S. Morasca and V. Basili, "An operational process for goal- driven definition of measures",. IEEE Transaction on Software Engineering , vol 30, no 2, pp 120-140, 2002.

AUTHORS PROFILE

Ritu Shrivastava has taught computer science to graduate students for 17 yrs in institutions like MANIT, Bhopal, Amity University, Delhi. She is actively involved in research in the field of object-oriented software engineering/technology.
e-mail ritushrivastava08@gmail.com

Dr J. L. Rana is retired professor of Computer Science & Engineering. He has 42 years experience of teaching and research. He has guided 6 candidates for Ph. D. degree and 2 are working under his guidance. His current research interests are Ad hoc Mobile Networks, Software Engineering.

e-mail jlrana@yahoo.com

Dr Mahendra Kumar is presently Prof. & Dean of Computer Science at S.I.R.T. , Bhopal. He was Professor and Head Computer applications at M.A.N.I.T., Bhopal. He has 42 years of teaching and research experience. He has published more than 90 papers in National and International journals. He has written two books and guided 12 candidates for Ph. D. degree and 3 more are currently working. His research interests are Software Engineering, Cross Language Information Retrieval, Text Mining, and Knowledge Management.

e-mail prof.mkumar@gmail.com

ISOR: Intelligent Secure On-Demand Routing Protocol

[1] Moitreyee Dasgupta, [2]Gaurav Sandhu
[1]Department of Computer Science and Engg., JSS Academy of Technical Education, Noida, New Delhi,
[2]Department of Computer Science and Engg.
GTBIT, New Delhi, India.
email: [1]helloruna@yahoo.com, [2] gauravgtbit@yahoo.in

3Usha Banerjee
[3]Department of Computer Science & Engg.
College of Engineering Roorkee, Roorkee, India.
email: [3]ushaban@gmail.com

***Abstract*— MANETs are highly vulnerable to attacks due to their inherent characteristics of the lack of infrastructure and complexity of wireless communication. Considerable improvements have been made towards providing ad hoc network security and existent solutions apply cryptography, intrusion detection systems or reputation systems. However, these conventional defense lines are inefficient to put all attacks and intrusions off. Our approach is to study the behavior of the AODV routing protocol in the presence of blackhole attacks, one of the major Denial-of Service attacks. In the first phase of this research, we provide the detailed simulation methodology of black hole attacks, and detail out the steps of creating a new routing protocol named as Intelligent Secure On-Demand Routing protocol (ISOR) using NS-2. In ISOR, an intelligent prevention scheme has been presented where every node will behave intelligently to prevent black hole attacks. Simulation studies show that compared to the original ad hoc on-demand distance vector (AODV) routing scheme, our proposed solution can verify 75% to 98% of the routes to the destination depending on the pause times at minimum delay in the networks.**

Keywords- Blackhole attacks, DoS Attacks, MANET, Security in MANET routing protocol

I. INTRODUCTION

Wireless Ad hoc networks are composed of autonomous nodes that are self- managed without any infrastructure. Ad hoc networks have a dynamic topology such that nodes can easily join or leave the network at any time. They have many potential applications, especially, in military and rescue areas such as connecting soldiers on the battlefield or establishing a new network in place of a network which collapsed after a disaster like an earthquake. Ad hoc networks are suitable for areas where it is not possible to set up a fixed infrastructure. The absence of any central coordinator or base station makes the routing process a complex one as compared to cellular networks. Hence the responsibilities of a routing protocol, include exchanging route information, finding a feasible path to a destination based upon criteria such as hop length, minimum power requirement and life time of wireless link, gathering information about path breaks, mending the broken paths and utilizing minimum bandwidth. Besides acting as a host, each node also acts as a node to discover a path and forward packets to the correct node in the network. All the proposed routing protocols [1] [2] [3] [4] [9] [11] [12] [14] are vulnerable to the denial-of –service attacks [5] [6]. Gaining access of the valid routes causes rushing attack [3]. An attacker can attract traffic towards certain destinations in the nodes under its control and cause the packet to be forwarded along a route that is non optimal or even non-existent. A pair of attacker nodes may create a wormhole [13] and shortcut their flows between each-other. Even being a part of the forwarding path, malicious nodes may selectively drop some or all the data packets [7]. A malicious node can correctly participate in route discovery phase but it may fail to correctly forward data packets. The security solution should also ensure that each node indeed forwards the packet according to its routing table. The black hole can be implemented in the network layer as well as in the MAC layer and as a result the entire network will be compromised. In this paper we propose an intelligent black hole attack prevention scheme to ensure reliable routing and data forwarding. In this scheme every node will behave in an intelligent manner and detect the corrupted node. Once detected, the node will be blacklisted for a definite period of time.

Simulations have been done using NS-2 (Network Simulator version 2) [7]. A new protocol has been added to the existing functionalities of NS-2 and black hole attacks have been simulated using this new protocol. After having implemented the new routing protocol which simulates a black hole, tests were performed on wireless networks to compare the network performance with and without black holes in the network. As expected, the throughput in the network deteriorated considerably in the presence of a black hole. Later in the paper, we have implemented our proposed solution to eliminate the effects of black hole and the results obtained were evaluated.

The rest of the paper is organized as follows. In section 2 we analyze various modes of attacks in ad hoc mobile networks. Section 3 presents a brief review of existing work. In section 4 we present a network attacking model based on black hole attack for AODV. The simulation of a black hole attack [7] and the proposed protocol ISOR is presented in section 5. In section 6 we analyze and discuss the results of simulation of the proposed ISOR protocol. In this section we also put forward a comparative study between the normal AODV

protocol and our proposed ISOR protocol. In this section we present a solution model for some countermeasures against black hole attacks. This section also deals with the performance evaluation of our routing protocol and a comparison with the existing AODV routing protocol.

II. DIFFERENT TYPES OF DOS ATTACKS

Security [5] is the primary challenge to ad hoc wireless networks because of its infrastructure-less features, resource constraints and dynamic topology changes. The security issue in MANET for group communication [7] is even more challenging because of the involvement of multiple senders and multiple receivers. DoS attacks [6] are hard to detect and easy to implement by an attacker as no hardware is required to do so. These are considered to be the most vulnerable category of attacks for network layer thus needs more attention. The entire network may fail in the presence of such an attack. Some common types of DoS attacks [10] [12] [13] [3] are discussed briefly:

- **Blackhole Attacks-** An attacker can drop received routing messages, instead of relaying them as the protocol requires, in order to reduce the quantity of routing information available to other nodes. This is called *blackhole attack* by Hu et al., and is a "passive" and simple way to perform a Denial of Service. The attack can be done selectively (drop routing packets for a specified destination, a packet every *n* packets, a packet every *t* seconds, or a randomly selected portion of the packets) or in bulk (drop all packets), and may have the effect of making the destination node unreachable or downgrade communication in the network.
- **Wormhole Attacks-**The *wormhole attack* [10] [13] is quite severe, and consists in recording traffic from one region of the network and replaying it in a different region. It is carried out by an intruder node *X* located within transmission range of legitimate nodes *A* and *B*, where *A* and *B* are not themselves within transmission range of each other. Intruder node *X* merely tunnels control traffic between *A* and *B* (and vice versa), without the modification presumed by the routing protocol – e.g. without stating its address as the source in the packets header – so that *X* is virtually invisible. This results in an extraneous inexistent *A* - *B* link which in fact is controlled by *X*. *X* can afterwards drop tunneled packets or break this link at will. Two intruder nodes *X* and *X′*, connected by a wireless or wired private medium, can also collude to create a longer (and more harmful) wormhole.
- **Jellyfish Attacks –** In this attack, the attacker obeys all the routing protocol specifications but, delays the packet forwarding process for a certain period of time, resulting in a high end-to-end delay. This attack is difficult to detect as packet drop in this case is negligible.
- **Rushing Attacks-** An offensive that can be carried out against on-demand routing protocols is the *rushing attack*. Typically, on-demand routing protocols state that nodes must forward only the first received Route Request from each route discovery; all further received route requests are ignored. This is done in order to reduce cluttering. The attack consists, for the adversary, in quickly forwarding its Route Request messages when a route discovery is initiated. If the Route Requests that first reach the target's neighbors are those of the attacker, then any discovered route includes the attacker.
- **Neighborhood Attacks-** An intermediate node records its ID in the packet before forwarding it to the next node. In this type of attack, an attacker simply forwards the packet without recording its ID in the packet. This makes two nodes that are not within the communication range of each other believe that they are neighbors (i.e., one hop away of each other), resulting in a disrupted route.

III. REVIEW WORK

Blackhole attack is one of the most active DoS attacks possible in MANETs. Research on black hole attacks has gained sufficient momentum. Research focuses mainly on securing existing routing protocols, developing new secure routing protocols, and intrusion detection techniques.

In [15] and [16] new protocols have been designed. Awerbuch et al. [15] developed a secure new on-demand routing protocol. It includes link weights which are considered during route discovery. The weights are calculated from the packet delivery fraction of each link. A link not delivering a fraction of packets above a certain threshold is considered malicious, and therefore the link weight is increased such that the link is chosen with smaller probability in the next route discovery phase. The approach detects a black hole as soon as the impact occurs, not when the black hole is constructed. In [16] a secure routing protocol based on the Dynamic Source Routing (DSR) protocol is presented. The authenticity of Route Requests is verified using message authentication codes (MAC). Furthermore, the authors present three techniques for authenticating data in Route Requests and Route Replies, where a broadcast authentication protocol for authenticating routing messages called TESLA ([17], [18]), digital signatures or MACs are used. Additionally, the authors propose per-hop hashing to verify that no node present in the node list of the Route Request is removed by an attacker. Finally, similar to the work done in [15] routes are chosen with regard to their prior performance in packet delivery. The work focuses on authentication of messages for on-demand protocols. Therefore, their approach is not applicable for pure ad hoc networks.

IV. BLACK-HOLE ATTACKING MODEL

Wireless Ad hoc networks are composed of autonomous nodes that are self- managed without any infrastructure. Besides acting as a host, each node also acts as a node to discover a path and forward packets to the correct node in the network. The AODV protocol is vulnerable to the well-known black hole attack. An attacker first introduces itself in the forwarding group (e.g., by implementing rushing attack), and then instead

This work is part of a WOS-A DST project (ref. no. SR/WOS-A/ET 20/2008, Department of Science & Technology, Government of India)

of forwarding the data packet to the proper destination, it simply drops all of data packets it receive resulting a poor packet delivery ratio [10].

In blackhole attack, the malicious node waits for the neighbors to initiate a RREQ packet. As soon as the malicious node receives the RREQ packet, it will immediately send a false RREP packet with a modified higher sequence number. Thus, the source node assumes that the node has a fresh route towards the destination. The source node ignores the RREP packet received from other nodes and begins to send the data packets to the malicious node. A malicious node takes all the routes towards itself. It does not allow forwarding any packet anywhere. This attack is called a blackhole as it swallows all objects and data packets [15].

A black hole is a node that always responds positively with a RREP message to every RREQ, even though it does not really have a valid route to the destination node. When the data packets routed by the source node reach the black hole node, it drops the packets rather than forwarding them to the destination node. The attacker may drop all data packets, or it may selectively drops the data packets. Discarding all data packets make the entire networks fail while selective dropping will degrade the network performance drastically.

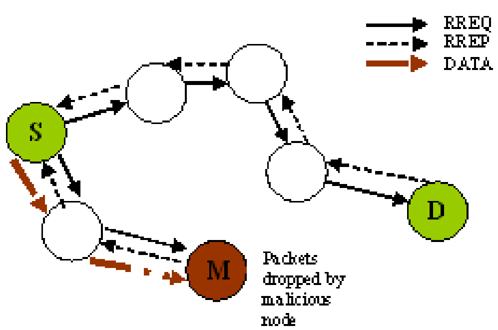


Figure 1: Blackhole attacking model

In figure 1, source node S wants to send data packets to a destination node D in the network. Node M is a malicious node which acts as a blackhole. The attacker replies with false reply RREP having higher modified sequence number. So, data communication initiates from S towards M instead of D.

V. ISOR: INTELLIGENT SECURE ON DEMAND ROUTING PROTOCOL

Sources broadcasts RREQ to the network for on-demand route discovery. On receiving the RREQ packet the malicious node immediately replies back with RREP packet with highest value of destination sequence number, ignoring the value in its routing table. Simultaneously, the destination node too replies with a RREP packet. After receiving the first RREP, the source does not choose that path, rather it adds that packet to a buffer linked list. The packet is also added to the suspicious linked list if its destination sequence number has a very large value. This process repeats itself a number of times to study the behavior of the suspicious nodes. The variable RREQ_COUNT has been used for this purpose. A decent simulation outcome has been obtained by setting the value of the variable equal to 10. There is a second count variable associated with each entry in the suspicious linked list which counts the number of times a highly acceptable (with larger sequence number) packet is send by a node. If the value of this count variable attains a value of RREQ_COUNT, it implies that a node is sending a RREP packet with a higher valued destination sequence number thus behaving in an ill manner. So, all the entries which follow the above pattern in the suspicious linked list are added to the black hole linked list. The suspicious node linked list gets destroyed once the nodes from the list get into the black hole linked list or after a decent amount of time. A black_final variable whose value is set to 50 is used which will destroy blackhole linked list when the RREQ packets count reaches to 50.

As the malicious node sends an RREP message without checking the tables, it is assumed that it is more likely for the first RREP message to arrive to the source. To nullify the attack, additional lists for suspicious nodes and Black Hole nodes were created and algorithms were applied to these linked lists to find the malicious node. These algorithms are explained in the next section.

The implementation requires two linked lists: SUSPICIOUS_NODE LINKED LIST and BLACKHOLE_NODE LINKED LIST. The discussion of the mechanism to counter Black Hole attack begins with the description of the two linked lists.

- SUSPICIOUS_NODE LINKED LIST: It is a linked list of the nodes which send the first RREP message, designed at requesting node. It contains the first RREP sending node's address, their destination sequence number with a count of how many times first RREP packets were sent by this node.
- BLACKHOLE_NODE LINKED LIST: It is a linked list which is created from the suspicious node linked list at the requesting node. It contains the RREP sending nodes address only. This list provides

V.1 ALGORITHM FOR ADDITION OF AN ENTRY INTO SUSPICIOUS NODE LINKED LIST

//ALGO FOR SUSPICIOUS LINKED LIST // SNLL

Assuming n1 is the node from where the first RREP has been received by the source;

nsaddr $\longleftarrow$ 0;
dst_seq $\longleftarrow$ 0; cnt $\longleftarrow$ 0; // global initialization;

do {
nsaddr $\longleftarrow$ address of n1;
dst_seq $\longleftarrow$ destination sequence of RREP received from n1;

If dst_seq $\cong$ 4294967290 then {
If ($\forall$ SNLL·nsaddr $\neq$ nsaddr) then

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

{Add the information to the suspicious node linked list;
cnt ⟵ 1 ;}
Else cnt ⟵ cnt + 1 ;}
while (recvReply);

V.2 Algorithm for destroying complete suspicious node linked list

If (SNLL⟶ start $\neq$ NULL) then {
Calculate the number of nodes in the list;
Destroy the lost ;}

Algorithm for addition of a node into Black Hole linked list:

// Algorithm for adding a node into Black hole Linked List // BHLL

do {
if (SNLL⟶nsaddr $\neq$ BHLL⟶nsaddr) then {
Add the node to BHLL;
Check for the next link ;}
} while SNLL⟶ next $\neq$ NULL;

V.3 Algorithm for filler function which copies black hole nodes from suspicious node linked list to black hole linked list

$\forall$(SNLL⟶nsaddr) if RREQ_COUNT = RREQ_FINAL then {
Add that node address to the Black Hole linked list. }

After designing the linked lists, a code which is written in the "*recvReply*" function of the AODV protocol tests whether the destination address of a packet is for that particular node or not.
Case 1 : If it is for the particular node-
The black hole node linked list is checked whether it is filled or not. If it is not, then that RREP packet is checked for the first reply. If it is the first reply then it is added to suspicious node linked list and "*recvReply*" function is exited.

Case 2: If black hole node list is present
If the black hole linked list contains node addresses, the list is checked for that particular node. If that node is present then the RREP message was from the black hole node and the "*recvReply*" function exits else that RREP information is checked for the first reply and is added to suspicious node linked list if true.

VI. Results and Analysis

Simulations are done using Network Simulator-2 [8].

Simulation Parameters

Radio: 802.11
Bandwidth: 2 Mbps
Nominal Range: 250 meters
Simulation time: 100 seconds
No. of node: 10
Mode of Placement: Random
Area: 800 meter by 800 meter
Placement of Malicious Node: Center
Connection between nodes: TCP
Traffic generation application: CBR (Constant Bit Rate)
Duration of the scenario: 10 seconds
Start time of CBR connections: First second of the scenario
End time of CBR connections: 100 seconds of the scenario
Packet Size: 1000 byte.
No. of repetitions: 20 times each for 0 and single malicious nodes for AODV.

In presence of the malicious nodes we have used ISOR and measured the performance of the network using packet delivery ratio as a performance metric.

Figure 2:
Figure 2 shows that normal AODV protocol is unable to function normally in presence of a black hole attack. There is no packet delivery in such a situation. The simulation results show that the entire network fails in presence of black hole attack as no connection is made from source to destination. As a result, no data packets were received by the destination node. Since there was a compromised node, even if other valid routes were available, the path containing malicious node only got selected because of high value sequence number of the RREP message send by the malicious node. After gaining access to the data forwarding path, malicious node dropped the entire data packet.

Figure 3:
Figure 3 shows the working of normal AODV protocol and our proposed ISOR protocol in normal conditions. There is no attack in this scenario and the graph verifies that both AODV and ISOR perform equally well.

Figure 4:
The main motivation for this research work stems from the fact that normal AODV protocol failed to work in a scenario which was infected with a black hole attack. Thus, we propose the ISOR protocol. This graph shows that with ISOR, the path has been established and data packets were received by the destination node. Figure 4 shows that in presence of a black hole attack when normal AODV failed. This graph proves that in such conditions of black hole attack AODV shows 100% data loss while ISOR worked fine. The results depicted in figure 4 prove that our proposed routing protocol, ISOR would verify 75% to 98% of the route to the destination depending on the pause times at the cost of minimum delay in the network.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

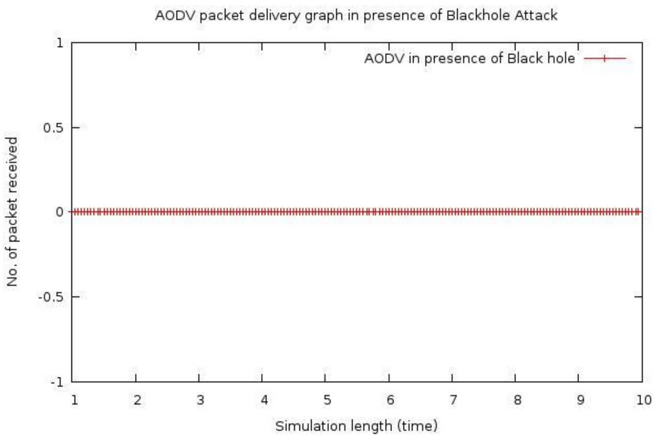


Figure 2: Graph showing AODV packet delivery in presence of Black Hole Attack

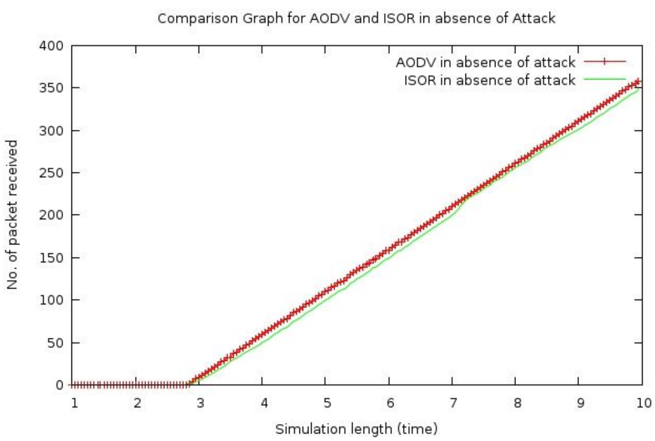


FIGURE 3: Graph showing comparison of AODV and ISOR in absence of attack

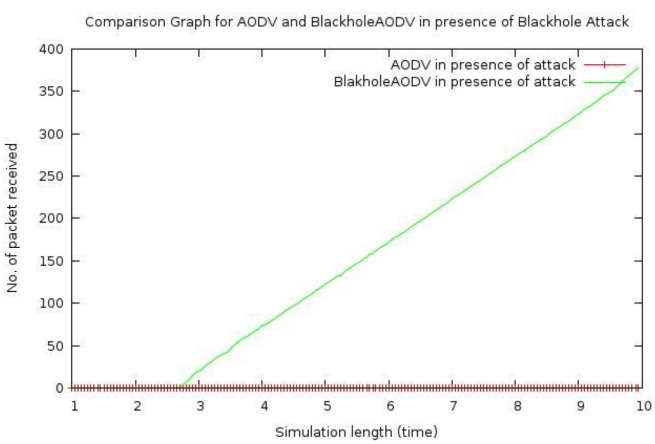


Figure 4: Graph showing Comparison of AODV in presence of Black Hole and ISOR

VII. CONCLUSION AND FUTURE WORK

In this paper, we have initially developed an attacking model for black hole attack. Later we have proposed a solution scheme to prevent black hole attack at the time of creation of black hole for unicast on-demand routing protocols. The attack is arrested at the time of creation of black hole and need not wait till it degrades the performance of the network. The proposed technique has the potential to prevent other types of viz. wormhole attacks as well. In fact, this algorithm can thwart all such attacks that need to get access to the packet forwarding group to carry out the attack. The biggest advantage of this algorithm is that it involves no extra overhead to the system. Malicious node would be blacklisted and an alternate path would be detected only if the sequence number value gets abnormally high due to some malicious behavior detected for the same route. We propose to extend this work presented in this paper to enable preventive measures using cross-layer information.

In this paper we have analyzed the behavior of the AODV protocol in presence of a black hole attack. However, other routing protocols could be simulated as well. All routing protocols are expected to present different results. Therefore, the best routing protocol for minimizing the effects of black hole attack may be determined. The algorithm that we have proposed makes communicating nodes intelligent.

REFERENCES

[1] M. Dasgupta, S. Choudhury and N. Chaki, “Secure Hypercube based team multicast routing protocol (S-HTMRP)”, Proceedings of First IEEE International Advanced Computing Conference (IACC’09), March 2009.

[2] Y Yi, M. Gerla, and K. Obraczka “Scalable Team Multicast Team in wireless networks exploiting coordinated motion”, Ad hoc Networks Journal, pp. 171-184, Aug 2003.

[3] Y. C. Hu, A. Perrig and D. B. Johnson “Rushing Attacks and Defense in Wireless Ad Hoc Networks Routing Protocol” Proceedings of ACM WiSe2003, Sep, 2003.

[4] C. E. Perkins and E. M. Royer, “Multicast ad hoc on-demand Distance Vector (MAODV) routing,” IETF draft, July 2001. http://www.ietf.org/proceedings/00dec/ID/draft-ietf-manet-maodv-00.txt

[5] H. Yang, H Y. Luo, F Ye, S W. Lu and L Zhang “Security in mobile ad hoc networks: Challenges and solutions” Proceedings of IEEE Wireless Communications, Pages 38-47, 2004.

[6] Hoang Lan Nguy en and Uyen Trang Nguyen “Study of Different Types of Attacks on Multicast in Mobile Ad Hoc Networks”, Proceedings of the International Conference of Networking, International Conference on Systems and International Conference on Mobile Communication and Learning Technologies.

[7] Tamilselvan, L. Sankaranarayanan, V. “Prevention of Blackhole Attack in MANET”, Proceedings of the 2nd International Conference on Wireless Broadband and Ultra Wideband Communications, 2007. (AusWireless 2007).

[8] “The network simulator - ns2,” http://www.isi.edu/nsnam/ns/.

[9] C. Perkins, E. Belding-Royer and S. Das “Ad hoc On-Demand Distance Vector (AODV) Routing” IETF draft, July 2001. Available: http://www.ietf.org/rfc/rfc3561.txt.

[10] Yih-Chun Hu, Adrian Perrig and David B. Johnson, “Packet Lashes: A Defense against Wormhole Attacks in Wireless Ad Hoc Networks” In Proceedings of the 22nd Annual Joint Conference of the IEEE Computer and Communication Societies (INFOCOM 2003), April 2003.

[11] M. G. Zapata “Secure Ad hoc On-Demand Distance Vector (SAODV) Routing” Available at http://personals.ac.upc.edu/guerrero/papers/draft-guerrero-manet-saodv-06.txt

[12] M. Gerla, G. Pei, S. J. Lee, C. C. Chiang “On Demand Multicast Routing Protocol for Mobile Ad Hoc Networks” Available at http://tools.ietf.org/html/draft-ietf-manet-odmrp-00

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[13] J. Eriksson, S. Krishnamurthy, and M. Faloutsos, “Truelink: A practical countermeasure to the wormhole attack in wireless networks,” in Proc. of ICNP’06. IEEE, 2006.

[14] R. Manoharan and P.Thambidurai “Hypercube Based Team Multicast Routing Protocol for Mobile Ad hoc Networks” Proceedings of 9th International Conference on Information Technology (ICIT’06).

[15] B. Awerbuch, D. Holmer, C. Nita-Rotaru, and H. Rubens, “An On-Demand Secure Routing Protocol Resilient to Byzantine Failures,” Proceedings of the 3rd ACM Workshop on Wireless Security, 2002.

[16] Y. Hu, A. Perrig, and D. Johnson, “Ariadne: A Secure On-Demand Routing Protocol for Ad Hoc Networks,” Proceedings of the 8th ACM International Conference on Mobile Computing and Networking, 2002.

[17] A. Perrig, R. Canetti, D. Song, and J. Tygar, “Efficient and Secure Source Authentication for Multicast,” In Network and Distributed System Security Symposium, pp. 35–46, February 2001.

[18] J. T. A. Perrig, R. Canetti and D. Song, “Efficient Authentication and Signing of Multicast Streams over Lossy Channels,” In IEEE Symposium on Security and Privacy, pp. 56–73, May 2000.

AUTHORS PROFILE

1. MOITREYEE DASGUPTA is an Assistant Professor in the Department of Computer Science and Engineering, JSS Academy of Technical Education, Noida, India. She had obtained her M.Tech degree in Computer Science & Engineering from the University of Calcutta, India in 2002. She has more than 9.5 years of experience in academics comprising of teaching and research. Her current research interests include mobile ad hoc networks, security in pure ad hoc networks and multicast routing protocols for ad hoc networks.

2. GAURAV SANDHU is a Lecturer in the Department of Information Technology at Guru Tegh Bahadur Institute of Technology, Delhi. He has a total work experience of 8 years in teaching Post-Graduate, Graduate and Engineering students. He holds a B.Tech degree in Electronics and Communication and M.Tech degree in Information Technology.

3. USHA BANERJEE is a Senior Lecturer in the Department of Computer Science and Engineering, College of Engineering Roorkee, Roorkee, India. She is also the Principal Investigator of a project sponsored by the Department of Science & Technology, Government of India. Her research interests are MANETs, Intrusion Detection Systems, Network security and performance of mobile networks. She graduated from Jadavpur University, Kolkata India in 2005 and obtained her M. Tech. degree with specialization in the area of Mobile Computing. She has an experience of 8 years in the IT industry and in academics

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

High Performance FingerPrint Identification System

Dr.R.Seshadri ,B.Tech,M.E,Ph.D
Director, S.V.U.Computer Center
S.V.University, Tirupati
E-mail : ravalaseshadri@gmail.com

Yaswanth Kumar.Avulapati,M.C.A,M.Tech,(Ph.D)
Research Scholar, Dept of Computer Science
S.V.University, Tirupati
E-mail:yaswanthkumar_1817@yahoo.co.in

Abstract

Biometrics is the science of establishing the identity of an individual based on their physical, chemical and behavioral characteristics of the person. Fingerprints are the most widely used biometric feature for person identification and verification in the field of biometric identification .A finger print is the representation of the epidermis of a finger. It consists of a pattern of interleaved ridges and valleys.

Fingerprints are graphical flow-like ridges present on human fingers. They are fully formed at about seven months of fetus development and finger ridge configurations do not change throughout the life of an individual except due to accidents such as bruises and cuts on the fingertips. This property makes fingerprints a very attractive biometric identifier. This paper presents an approach to classifying the fingerprints into different groups and increase the performance of the system.It increases the performance of fingerprint matching while matching the input template with stored template.

Keywords-Biometrics, Verification, Identification

Introduction

A fingerprint is a pattern of ridges and valleys located on the tip of each finger. Fingerprints were used for personal identification for many centuries and the matching accuracy was very high. Patterns have been extracted by creating an inked impression of the fingertip on paper.

Today, compact sensors provide digital images of these patterns. Fingerprint system can be separated into two categories Verification and identification.

Verification system authenticates a person’s identity by comparing the captured biometric characteristic with its own biometric template(s) pre-stored in the system It conducts one-to-one comparison to determine whether the identity claimed by the individual is true.

A verification system either rejects or accepts the submitted claim of identity Identification system recognizes an individual by searching the entire template database for a match. It conducts one-to-many comparisons to establish the identity of the individual.

In an identification system, the system establishes a subject’s identity without the subject having to claim an identity.

Prehistoric picture writing of a hand with ridge patterns was discovered in Nova Scotia. In ancient Babylon,fingerprints were used on clay tablets for business transactions. In ancient China, thumb prints were found on clay seals.In 14th century Persia, various official government papers had fingerprints (impressions), and one government official, a doctor, observed that no two fingerprints were exactly alike.

In 1686, Marcello Malpighi, a professor of anatomy at the University of Bologna, noted in his treatise; ridges, spirals and loops in fingerprints. He made no mention of their value as a tool for individual identification. A layer of skin was named after him; "Malpighi" layer, which is approximately 1.8mm thick.

In 1823, John Evangelist Purkinji, a professor of anatomy at the University of Breslau, published his thesis discussing 9 fingerprint patterns, but he too made no mention of the value of fingerprints for personal identification. During the 1870's, Dr. Henry Faulds, the British Surgeon-Superintendent of Tsukiji Hospital in Tokyo, Japan, took up the study of "skin-furrows" after noticing finger marks on specimens of "prehistoric"

pottery. A learned and industrious man, Dr. Faulds not only recognized the importance of fingerprints as a means of identification, but devised a method of classification as well.

In 1880, Faulds forwarded an explanation of his classification system and a sample of the forms he had designed for recording inked impressions, to Sir Charles Darwin. Darwin, in advanced age and ill health, informed Dr. Faulds that he could be of no assistance to him, but promised to pass the materials on to his cousin, Francis Galton.

Also in 1880, Dr. Faulds published an article in the Scientific Journal, "Nautre" (nature). He discussed fingerprints as a means of personal identification, and the use of printers ink as a method for obtaining such fingerprints. He is also credited with the first fingerprint identification of a greasy fingerprint left on an alcohol bottle.

In order to implement a fingerprint system, the various research methodologies involved in it like fingerprint image capture, image preprocessing, feature extraction, storage and image matching must be clearly defined are shown in figure. 1

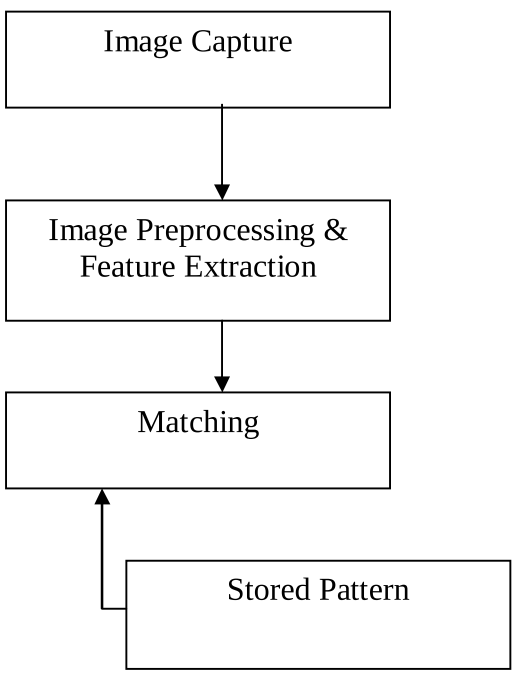


Fig 1.Various steps in a Fingerprint system

A finger print system works in two different modes they are **Enrollment mode** and **Authentication mode** as shown in figure.2. Enrollment mode in which fingerprint system is used to identify and collect the related information about the person and his/her fingerprint image.
Authentication mode in which fingerprint system is used to identify the person who is declared to be him/her.

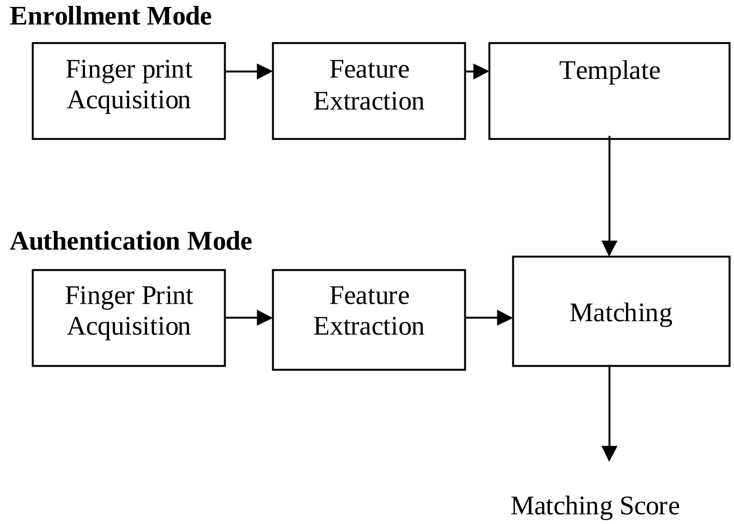


Fig 2.Enrollment and Authentication of a fingerprint system

Fingerprint matching can be performed based on Minutiae, Correlation based, Ridge feature based. In minutiae based matching it stores minutiae is a set of points in a plane and the points are matched in the template and the input minutiae. In correlation based matching two finger print images and correlation between corresponding pixels is computed. Ridge feature based is a advanced technology that capture the ridges. The most popular technologies used to identify fingerprint are Optical, silicon and ultra sound.

Previous Work:

The previous work is based on the theory of fingerprint classification they stored only single finger print of person in the database. This single finger print can be index or thumb. Let us see how the previous system will work. In the enrollment process in conventional system the database contains the fingerprint templates in an ordinary manner but in that system the database e contains the different set of templates according to classification. During the enrollment process, sensor sense the fingerprint, then next step is feature extraction . After this step they put a classifier to check the classification of input template that whether it is left-loop, right-loop, arch or whorl as shown in the figure 3

Let us come to the verification process here the finger print is placed at sensor and then its features are extracted and a final template is generated for matching. Now this template will not matched with every templates in the database rather it extracts its classified domain out of 4-domain and will perform match from this extracted domain

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

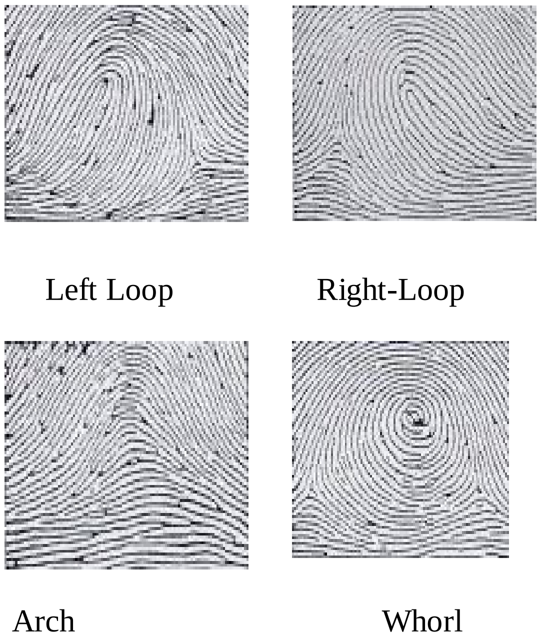


Fig3.Classification of Finger prints in existing system

Fingerprint classifiers classify the input fingerprint into four major categories namely Left-Loop, Right-Loop, Whorl and Arch. They proposed classifiers works on the basis of singular point (Delta) extracted. If there are two deltas then it will be counted as whorl or twin loop. If there is no delta then it will be counted as arch. If only one delta is there then it will be consider as either left loop or right loop.

Problems in the Existing system:

The existing system can identify the finger prints according to their four categories namely Left-Loop, Right-Loop, Whorl and Arch.

If the people having different types of finger prints other than this four categories. It is very difficult for the system to identify the finger prints like mixed category, pocked loop, double loop. The time taken for identifies the finger prints is also more in the existing system. It decreases the performance of the system.

Proposed Work:

Proposed work is based on the classification of fingerprints. In our proposed system during the enrollment process fingerprint is captured with a sensor, then next step is feature extraction . we further classify the finger prints as arch,tentarch,loop, doubleLoop, pocked Loop, whorl ,mixed, left-loop, right-loop as shown in figure below

After this step we put a classifier to check the classification of input template that whether it is arch,tentarch, loop, doubleLoop, pocked Loop, whorl ,mixed, left-loop, right-loop

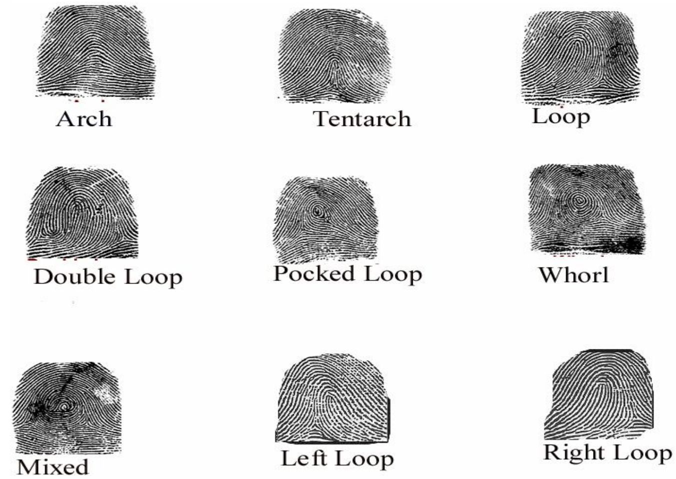


Fig4.Classification of Finger prints in proposed system

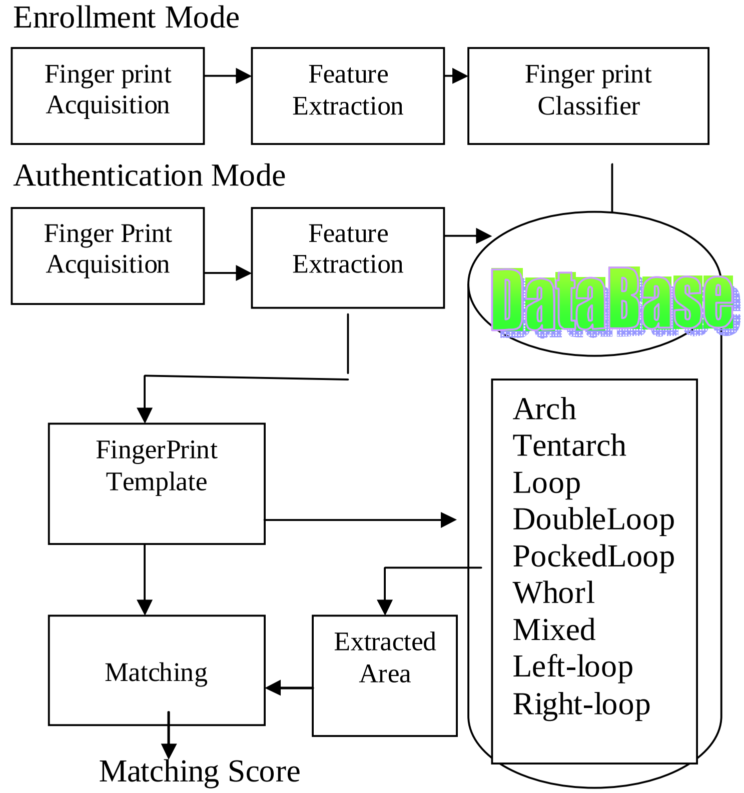


Fig5.Proposed Fingerprint identification system

After classification the input template is stored in particular area. A area in the database contains the templates of same classification. Normally fingerprints are classified as Whorl(27%), arch (4%) loops(65%)and mixed (4%) we further divide this domain into four parts i.e. left loop (26%) right loop (25%) pocked loop (9%)and double loop (5%), apx

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Fingerprint Classifier:

The proposed classifiers works on the basis of core and Delta extracted. If there is two deltas then it will be counted as whorl or twin loop. If there is no delta then it will be counted at arch. If only one delta is there then it will be either left loop or right loop. If there is only one delta and one core then it will be pocked loop. If there is two deltas and one core then it will be double loop. If there is two deltas and two cores then it will be mixed as shown in the figure.6

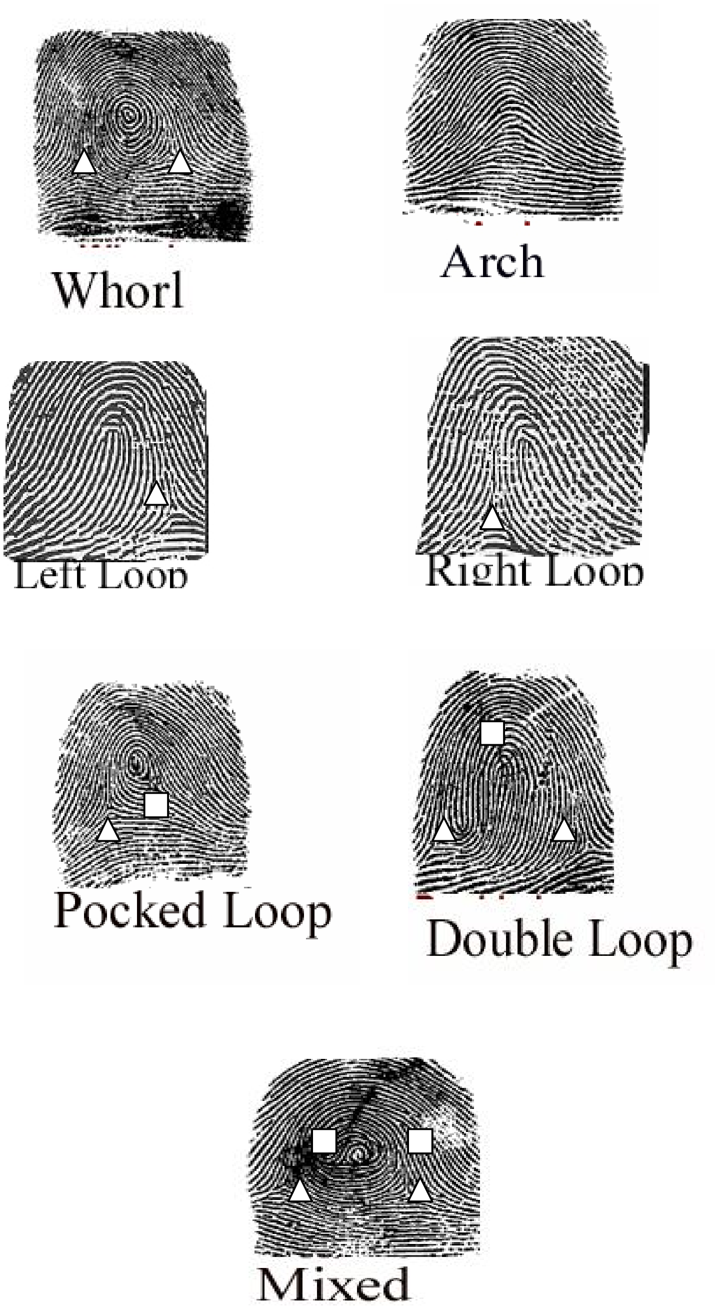


Fig 6. Position and number of Core and Delta in different Finger prints

Performance of Existing System

For Best case i.e. the template is First match, Time required = 1 X 1 = 1 ms they calculate for worst case They assumed 1, 50000 templates , According to c lassific ation there will be 45000 whorls (30%) + 48000 Left Loop (32%) + 49500 Right Loop (33%) + 7500 Arch (5%)

At First stage they get the template classification and accordingly particular domain will be extracted. Now they calculate the time taken for each classification
For Whorl = 1ms X 45000 = 45 sec.
For LL = 1ms X 48000 = 48 sec.
For RL = 1ms X 49500= 49.5 sec
For Arch = 1msX 7500= 7.5 sec.

Average time = 150/4= 37.4 sec.
For an Average case, Time required= apx 20-24 sec.

Proposed System Fingerprint classification:

Let us assume that we classify fingerprints as Whorl, loop,mixed. Loops make up nearly 65% of all fingerprints, whorls are nearly 27%, arches are nearly 4%and mixed are nearly 4% Since the loops are 65%, we further divide this domain into four parts i.e. left loop 26% right loop 25% pocked loop 9%and double loop 5%, apx .

Performance of Proposed System

For Best case i.e. the template in First match, Time required = 1 X 1 = 1 msNow let us calculate for worst case We have assumed 1, 50000 templates , According to classific ation there will be 40500whorls (27%) + 39000 Left Loop (26%) + 37500 Right Loop (25%) + 13500 Pocked Loop (9%) +7500 Double Loop(5%)+6000 Arch(4%)+6000 Mixed(4%).At First stage we get the template classification and accordingly particular domain will be extracted. Now we calculate the time taken for each classification
For Whorl = 1ms X 40500 = 40..5 sec.
For LL = 1ms X 39000 = 39 sec.
For RL = 1ms X 37500= 37.5 sec
For PL = 1ms X 13500= 13.5 sec
For DL = 1ms X 7500= 7.5 sec
For Mixed = 1ms X 6000= 6 sec
For Arch = 1msX 6000= 6 sec.

Average time = 150/7= 21.42sec.
For an Average case, Time required= apx 12-18 sec.

Performance Factor

PF=Time taken in worst case of existing system 37.4sec

PF=Time take in worst case of proposed system 21.42 Sec

i.e. the new approach is better than the existing one.

Coclusion:

This paper presents an approach to classifying the fingerprints into different groups and increase the performance of the system. It increases the performance of fingerprint matching while matching the input template with stored template.The paper presents an overview of the different steps involved in the enrollment and authentication modes. We have proposed seven major classifications of fingerprints like arch, doubleLoop, pocked Loop, whorl ,mixed, left-loop, right-loop .Its a new approach for classification of fingerprints and matching in the database. This paper presents an approach to speed up the matching process by classifying the fingerprint into different groups at the time of enrollment, and authentication modes. This proposed system is better than the previous one.

References:

1.A. K. Jain,Patrick Flynn,Arun A.Ross . "Handbook of Biometrics".
2.D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar, Handbook of Fingerprint Recognition. Springer, 2003.
3. C. D. Kuglin and D. C. Hines, "The phase correlation image alignment method," Proc. Int. Conf. on Cybernetics and Society, pp. 163–165, 1975.
4. K. Takita, T. Aoki, Y. Sasaki, T. Higuchi, and K. Kobayashi, "High-accuracy subpixel image registration based on phase-only correlation," IEICE Trans. Fundamentals,vol. E86-A, no. 8, pp. 1925–1934, Aug. 2003.
5. K. Takita, M. A. Muquit, T. Aoki, and T. Higuchi, "A subpixel correspondence search technique for computer vision applications," IEICE Trans. Fundamentals, vol. E87-A, no. 8, pp. 1913–1923, Aug. 2004.

6. K. Ito, H. Nakajima, K. Kobayashi, T. Aoki, and T. Higuchi, "A fingerprint matching algorithm using phaseonlycorrelation," IEICE Trans. Fundamentals, vol. E87-A,no. 3, pp. 682–691, Mar. 2004.
7. M. Kawagoe and A. Tojo, "Fingerprint pattern classification," Pattern Recognition, vol. 17, no. 3, pp. 295–303, 1984.
8. www.aladdinusa.com
Biometrics Information Group www.biometricsinfo.org.
9. A. K. Jain, R. Bolle, S. Pankanti (eds), Biometrics: Personal Identification in Networked Society, Kluwer Academic, December 1998.
10.Reducing Process-Time for Fingerprint Identification System , Chander Kant & Rajender Nath
11. A. K. Hrechak and J. A. McHugh, Automated Fingerprint Recognition using Structural Matching, Pattern Recognition, Vol. 23, No. 8, 1990.
12. The Henry Clas sification System Copyright © 2003 International Biometric Group
13.A. K. Jain, S. Prabhak ar. "Handbook of Fingerprint Recognition".
14.www.google.co.in
15A. K. Jain and S.Pankanti , "Fingerprint lassification and matching, "InA. Bovik, Ed., Handbook for Image and Video Processing . AcademicPress, 2000.
16. Z.W. bo, N. X. bao and W.C. jian, " A fingerprint matching algorithmbased on relative topological relationship among minutiae, " IEEE Int. Conference Neural Networks & Signal Processing Zhenjiang, China,2008.
17. Y.He, J.Tian, X.Luo and T.Zhang, " Image enhancement and minutiae matching in fingerprint verification, "Patt.Recog.Lett. no.24, 2003,pp.1349-1360.
18.Glossary of Biometric Terms, Association for Biometrics and International Computer Security Association, to be referred at URL: http://www.afb.org.uk/ (1998).
19. Biometrics Information Management and Security (2001).

20. Bahuguna, R.D. and Corboline, T.: Prism fingerprint sensor that uses a holographic optical element, *APPLIED OPTICS*, Vol. 35, No. 26 (1996).

Authors Profile

Dr.R.Seshadri was born in Andhra Pradesh, India, in 1959. He received his **B.Tech** degree from Nagarjuna University in 1981. He received his **M.E** degree in Control System Engineering from PSG College of Technology, Coimbatore in 1984. He was awarded with **PhD** from Sri Venkateswara University, Tirupati in 1998. He is currently Director, Computer Center, S.V.University, Tirupati, India. He has Published number of papers in national and international conferences, seminars and journals. At present 12 members are doing research work under his guidance in different areas

Mr.YaswanthKumar .Avulapati received his **MCA** degree with **First class** from Sri Venkateswara University, Tirupati. He received his **M.Tech** Computer Science and Engineering degree with **Distinction** from Acharya Nagarjuna University, Guntur.He is a research scholar in S.V.University Tirupati, Andhra Pradesh.He has presented number of papers in national and international conferences, seminars.He attend Number of work shops in different fields.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Constraint-free Optimal Meta Similarity Clusters Using Dynamic Minimum Spanning Tree

S. John Peter
Assistant Professor
Department of Computer Science and Research Center
St. Xavier's College, Palayamkottai
Tamil Nadu, India.
jaypeeyes@rediffmail.com

S.P. Victor
Associate Professor
Department of Computer Science and Research Center
St. Xavier's College, Palayamkottai
Tamil Nadu, India.
victorsp@rediffmail.com

ABSTRACT — **Clustering is a process of discovering groups of objects such that the objects of the same group are similar, and objects belonging to different groups are dissimilar. A number of clustering algorithms exist that can solve the problem of clustering, but most of them are very sensitive to their input parameters. Therefore it is very important to evaluate the result of them. The minimum spanning tree clustering algorithm is capable of detecting clusters with irregular boundaries. In this paper we propose a constraint-free minimum spanning tree based clustering algorithm. The algorithm constructs hierarchy from top to bottom. At each hierarchical level, it optimizes the number of cluster, from which the proper hierarchical structure of underlying dataset can be found. The algorithm uses a new cluster validation criterion based on the geometric property of data partition of the data set in order to find the proper number of clusters at each level. The algorithm works in two phases. The first phase of the algorithm create clusters with guaranteed intra-cluster similarity, where as the second phase of the algorithm create dendrogram using the clusters as objects with guaranteed inter-cluster similarity. The first phase of the algorithm uses divisive approach, where as the second phase uses agglomerative approach. In this paper we used both the approaches in the algorithm to find Optimal Meta similarity clusters.**

Keywords: Euclidean minimum spanning tree, Subtree, Clustering, Eccentricity, Center, Hierarchical clustering, Dendrogram, Cluster validity, Cluster Separation

I. INTRODUCTION

The problem of determining the correct number of clusters in a data set is perhaps the most difficult and ambiguous part of cluster analysis. The "true" number of clusters depends on the "level" on is viewing the data. Another problem is due to the methods that may yield the "correct" number of clusters for a "bad" classification [10]. Furthermore, it has been emphasized that mechanical methods for determining the optimal number of clusters should not ignore that the fact that the overall clustering process has an unsupervised nature and its fundamental objective is to uncover the unknown structure of a data set, not to impose one. For these reasons, one should be well aware about the explicit and implicit assumptions underlying the actual clustering procedure before the number of clusters can be reliably estimated or, otherwise the initial objective of the process may be lost. As a solution for this, Hardy [10] recommends that the determination of optimal number of clusters should be made by using several different clustering methods that together produce more information about the data. By forcing a structure to a data set, the important and surprising facts about the data will likely remain uncovered.

In some applications the number of clusters is not a problem, because it is predetermined by the context [11]. Then the goal is to obtain a mechanical partition for a particular data using a fixed number of clusters. Such a process is not intended for inspecting new and unexpected facts

arising from the data. Hence, splitting up a homogeneous data set in a "fair" way is much more straightforward problem when compared to the analysis of hidden structures from heterogeneous data set. The clustering algorithms [15, 21] partitioning the data set in to *k* clusters without knowing the homogeneity of groups. Hence the principal goal of these clustering problems is not to uncover novel or interesting facts about data.

Numerical methods can usually provide only guidance about the true number of clusters and the final decision is often an ad hoc decision that is based on prior assumptions and domain knowledge. Therefore, the choice between the different numbers of clusters is often made by comparing several alternatives, and the final decision is a subjective problem that can be solved in practice only by humans. Nevertheless, a number of methods for objective assessment of cluster validity have been developed and proposed. Because the recognition of cluster structures is difficult especially in high-dimensional spaces, various visualization technique can also be of valuable help to the cluster analyst.

Given a connected, undirected graph $G = (V, E)$, where V is the set of nodes, E is the set of edges between pairs of nodes, and a weight $w(u, v)$ specifying weight of the edge (u, v) for each edge $(u, v) \in E$. A spanning tree is an acyclic subgraph of a graph G, which contains all vertices from G. The Minimum Spanning Tree (**MST**) of a weighted graph is minimum weight spanning tree of that graph. Several well established **MST** algorithms exist to solve minimum spanning tree problem [24, 19, 20]. The cost of constructing a minimum spanning tree is $O(m \log n)$, where m is the number of edges in the graph and n is the number of vertices. More efficient algorithm for constructing **MST**s have also been extensively researched [18, 5, 13]. These algorithms promise close to linear time complexity under different assumptions. A Euclidean minimum spanning tree (**EMST**) is a spanning tree of a set of n points in a metric space ($\mathbf{E}^n$), where the length of an edge is the Euclidean distance between a pair of points in the point set.

The hierarchical clustering approaches are related to graph theoretic clustering. Clustering algorithms using minimal spanning tree takes the advantage of **MST**. The **MST** ignores many possible connections between the data patterns, so the cost of clustering can be decreased. The **MST** based clustering algorithm is known to be capable of detecting clusters with various shapes and size [34]. Unlike traditional clustering algorithms, the **MST** clustering algorithm does not assume a spherical shapes structure of the underlying data. The **EMST** clustering algorithm [23,34] uses the Euclidean minimum spanning tree of a graph to produce the structure of point clusters in the *n*-dimensional Euclidean space. Clusters are detected to achieve some measures of optimality, such as minimum intra-cluster distance or maximum inter-cluster distance [2]. The **EMST** algorithm has been widely used in practice.

Clustering by minimal spanning tree can be viewed as a hierarchical clustering algorithm which follows a divisive approach. Using this method firstly **MST** is constructed for a given input. There are different methods to produce group of clusters. If the number of clusters *k* is given in advance, the simplest way to obtain *k* clusters is to sort the edges of minimum spanning tree in descending order of their weights and remove edges with first *k*-1 heaviest weights [2, 33].

All existing clustering Algorithm require a number of parameters as their inputs and these parameters can significantly affect the cluster quality. Our algorithm does not require a predefined cluster number. In this paper we want to avoid experimental methods and advocate the idea of need-specific as opposed to care-specific because users always know the needs of their applications. We believe it is a good idea to allow users to define their desired similarity within a cluster and allow them to have some flexibility to adjust the similarity if the adjustment is needed. Our Algorithm produces clusters of *n*-dimensional points with a naturally approximate intra-cluster distance.

Geometric notion of centrality are closely linked to facility location problem. The distance matrix D can computed rather efficiently using Dijkstra's algorithm with time complexity $O(|V|^2 \ln |V|)$ [29].

The *eccentricity* of a vertex x in G and radius ρ (G), respectively are defined as

$$e(x) = \max_{y \in V} d(x, y) \quad \text{and} \quad \rho(G) = \min_{x \in V} e(x)$$

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

The *center* of *G* is the set

$$C(G) = \{x \in V \mid e(x) = \rho(G)\}$$

C (G) is the center to the "*emergency facility location problem*" which is always contain single block of *G*. The length of the longest path in the graph is called *diameter* of the graph *G*. we can define diameter D (G) as

$$D(G) = \max_{x \in V} e(x)$$

The *diameter* set of *G* is

$$Dia(G) = \{x \in V \mid e(x) = D(G)\}$$

An important objective of hierarchical cluster analysis is to provide picture of data that can easily be interpreted. A picture of a hierarchical clustering is much easier for a human being to comprehend than is a list of abstract symbols. A *dendrogram* is a special type of tree structure that provides a convenient way to represent hierarchical clustering. A dendrogram consists of layers of nodes, each representing a cluster.

Hierarchical clustering is a sequence of partitions in which each partition is nested into the next in sequence. An Agglomerative algorithm for hierarchical clustering starts with disjoint clustering, which places each of the *n* objects in an individual cluster [1]. The hierarchical clustering algorithm being employed dictates how the proximity matrix or proximity graph should be interpreted to merge two or more of these trivial clusters, thus nesting the trivial clusters into second partition. The process is repeated to form a sequence of nested clustering in which the number of clusters decreases as a sequence progress until single cluster containing all *n* objects, called the *conjoint clustering*, remains[1].

Nearly all hierarchical clustering techniques that include the tree structure have two short comings: (1) they do not properly represent hierarchical relationship and (2) once the data are assigned improperly to a given cluster it cannot later reevaluate and placed in another cluster.

In this paper, we propose a new clustering algorithm: the Dynamically Growing Minimum Spanning Tree (**DGMST**), which can overcome these shortcomings. The algorithm optimizes the number of clusters at each hierarchical level with the cluster validation criteria during the minimum spanning tree construction process. Then the hierarchy constructed by the algorithm can properly represent the hierarchical structure of the underlying dataset, which improves the accuracy of the final clustering result.

Our **DGMST** clustering algorithm addresses the issues of undesired clustering structure and unnecessary large number of clusters. Our algorithm does not require a predefined cluster number. The algorithm constructs an **EMST** of a point set and removes the inconsistent edges that satisfy the inconsistence measure. The process is repeated to create a hierarchy of clusters until optimal numbers of clusters (regions) are obtained. Hence the title! In section 2 we review some of the existing works on cluster validity and graph based clustering algorithms. In Section 3 we propose **DGMST** algorithm which produces optimal number of clusters with dendrogram for cluster of clusters. Hence we named this new cluster as *Optimal Meta similarity clusters.* Finally in conclusion we summarize the strength of our methods and possible improvements.

II. RELATED WORK.

Determining the true number of clusters, also known as the cluster validation problem, is a fundamental problem in cluster analysis. Many approaches to this problem have been proposed [25, 32, 10]. Two kinds of indexes have been used to validate the clustering [6, 7]: one based on relative criteria and other based on external and internal criteria. The first approach is to choose the best result from set of clustering result according to a prespecified criterion. Although the computational cost of the approach is light, human intervention is required to find the best number of clusters. The **DGMST** algorithm tries to find the proper number of clusters automatically which makes the first approach unsuitable for clustering validation in the **DGMST** algorithm**.**

The second approach is based on statistical tests and involves computations of both inter-cluster and intra-cluster quality to determine the proper best number of clusters. The evaluation of the criteria can be completed automatically. However the computational cost of this type of cluster validation is very high. The second type of this kind of approach is also not suitable for **DGMST** algorithm when it is used to cluster a large dataset. A successful and practical cluster validation criteria used in the **DGMST** algorithm

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

for large dataset must have modest computational cost and can be easily evaluated automatically.

Clustering by minimal spanning tree can be viewed as a hierarchical clustering algorithm which follows the divisive approach. Clustering Algorithm based on minimum and maximum spanning tree were extensively studied. Avis [3] found an $O(n^2 \log^2 n)$ algorithm for the min-max diameter-2 clustering problem. Asano, Bhattacharya, Keil and Yao [2] later gave optimal O $(n \log n)$ algorithm using maximum spanning trees for minimizing the maximum diameter of a bipartition. The problem becomes NP-complete when the number of partitions is beyond two [17]. Asano, Bhattacharya, Keil and Yao also considered the clustering problem in which the goal to maximize the minimum inter-cluster distance. They gave a *k*-partition of point set removing the *k*-1 longest edges from the minimum spanning tree constructed from that point set [2]. The identification of inconsistent edges causes problem in the **MST** clustering algorithm. There exist numerous ways to divide clusters successively, but there is not suitable a suitable choice for all cases.

Zahn [34] proposes to construct **MST** of point set and delete inconsistent edges – the edges, whose weights are significantly larger than the average weight of the nearby edges in the tree. Zahn's inconsistent measure is defined as follows. Let *e* denote an edge in the **MST** of the point set, v_1 and v_2 be the end nodes of *e, w* be the weight of *e*. A *depth neighborhood* ***N*** of an end node *v* of an edge *e* defined as a set of all edges that belong to all the path of length *d* originating from *v*, excluding the path that include the edge *e*. Let N_1 and N_2 be the depth *d* neighborhood of the node v_1 and v_2. Let $\hat{W}_{N1}$ be the average weight of edges in N_1 and σN_1 be its standard deviation. Similarly, let $\hat{W}_{N2}$ be the average weight of edges in N_2 and σN_2 be its standard deviation. The inconsistency measure requires one of the three conditions hold:

1. $w > \hat{W}N_1 + c \times \sigma N_1$ or $w > \hat{W}N_2 + c \times \sigma N_2$

2. $w > \max(\hat{W}N_1 + c \times \sigma N_1 , \hat{W}N_2 + c \times \sigma N_2)$

3. $$\frac{w}{\max(c \times \sigma N_1 , c \times \sigma N_2)} > f$$

where *c* and *f* are preset constants. All the edges of a tree that satisfy the inconsistency measure are considered inconsistent and are removed from the tree. This result in set of disjoint subtrees each represents a separate cluster. Paivinen [22] proposed a Scale Free Minimum Spanning Tree **(SFMST)** clustering algorithm which constructs scale free networks and outputs clusters containing highly connected vertices and those connected to them.

The **MST** clustering algorithm has been widely used in practice. Xu (Ying), Olman and Xu (Dong) [33] use MST as multidimensional gene expression data. They point out that **MST**- based clustering algorithm does not assume that data points are grouped around centers or separated by regular geometric curve. Thus the shape of the cluster boundary has little impact on the performance of the algorithm. They described three objective functions and the corresponding cluster algorithm for computing *k*-partition of spanning tree for predefined $k > 0$. The algorithm simply removes *k*-1 longest edges so that the weight of the subtrees is minimized. The second objective function is defined to minimize the total distance between the center and each data point in the cluster. The algorithm removes first *k*-1 edges from the tree, which creates a *k*-partitions.

The clustering algorithm proposed by S.C.Johnson [16] uses proximity matrix as input data. The algorithm is an agglomerative scheme that erases rows and columns in the proximity matrix as old clusters are merged into new ones. The algorithm is simplified by assuming no ties in the proximity matrix. Graph based algorithm was proposed by Hubert [12] using single link and complete link methods. He used threshold graph for formation of hierarchical clustering. An algorithm for single-link hierarchical clustering begins with the minimum spanning tree (MST) for G (∞), which is a proximity graph containing $n(n-1)/2$ edge was proposed by Gower and Ross [14]. Later Hansen and DeLattre [9] proposed another hierarchical algorithm from graph coloring.

Many different methods for determining the number of clusters have been developed. Hierarchical clustering methods provide direct information about the number of clusters by clustering objects on a number of different hierarchical levels, which are then presented by a graphical tree structure known as *dendrogram.* One may apply some external criteria to validate the solutions on different levels or use the dendrogram visualization for determining the best cluster structure.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

The procedure of evaluating the results of a clustering algorithm is known under the term cluster validity. In general terms, there are three approaches to investigate cluster validity [31]. The first is based on *external criteria*. This implies that we evaluate the results of a clustering algorithm based on a pre-specified structure, which is imposed on a data set and reflects our intuition about the clustering structure of the data set. The second structure is based on *internal criteria*. In this case the clustering results are evaluated in terms of quantities that involve the vectors of the data set themselves (e.g. proximity matrix). The third approach of clustering validity is based on *relative criteria*. Here the basic idea is the evaluation of a clustering structure by comparing it to other clustering schemes, resulting by the same algorithm but with different input parameter values.

The selection of the correct number of clusters is actually a kind of validation problem. A large number of clusters provides a more complex "model" where as a small number may approximate data too much. Hence, several methods and indices have been developed for the problem of cluster validation and selection of the number of clusters [27, 8, 26, 28, 30]. Many of them based on the within and between-group distance.

III. OUR CLUSTERING ALGORITHM

A tree is a simple structure for representing binary relationship, and any connected components of tree is called *subtree*. Through this **MST** representation, we can convert a multi-dimensional clustering problem to a tree partitioning problem, ie finding particular set of tree edges and then cutting them. Representing a set of multi-dimensional data points as simple tree structure will clearly lose some of the inter data relationship. However many clustering algorithm proved that no essential information is lost for the purpose of clustering. This is achieved through rigorous proof that each cluster corresponds to one subtree, which does not overlap the representing subtree of any other cluster. Clustering problem is equivalent to a problem of identifying these subtrees through solving a tree partitioning problem. The inherent cluster structure of a point set in a metric space is closely related to how objects or concepts are embedded in the point set. In practice, the approximate number of embedded objects can sometimes be acquired with the help of domain experts. Other times this information is hidden and unavailable to the clustering algorithm. In this section we present clustering algorithm which produce optimal number of clusters.

A. DGMST Clustering Algorithm:

Given a point set S in $\mathbf{E^n}$, the hierarchical method starts by constructing a Minimum Spanning Tree (**MST)** from the points in S. The weight of the edge in the tree is Euclidean distance between the two end points. So we named this **MST** as **EMST1.** Next the average weight $\hat{W}$ of the edges in the entire **EMST1** and its standard deviation σ are computed; any edge with $W > \hat{W} + \sigma$ or *current longest edge* is removed from the tree. This leads to a set of disjoint subtrees $S_T = \{T_1, T_2 \ldots\}$ *(divisive approach)*. Each of these subtrees T_i is treated as cluster. Oleksandr Grygorash et al proposed minimum spanning tree based clustering algorithm [21] which generates k clusters. Our previous algorithm [15] generates k clusters with centers, which used to produce Meta similarity clusters. Both of the minimum spanning tree based algorithms assumed the desired number of clusters in advance. In practice, determining the number of clusters is often coupled with discovering cluster structure. Hence we propose a new algorithm named, *Dynamically Growing Minimum Spanning Tree algorithm (***DGMST**), which does not require a predefined cluster number. The algorithm works in two phases. The first phase of the algorithm partitioned the **EMST1** into sub trees (clusters/regions). The centers of clusters or regions are identified using eccentricity of points. These points are a representative point for the each subtree S_T. A point c_i is assigned to a cluster i if $c_i \in T_i$. The group of center points is represented as $C = \{c_1, c_2 \ldots \ldots c_k\}$. These center points $c_1, c_2 \ldots . c_k$ are connected and again minimum spanning tree **EMST2** is constructed is shown in the Figure 4. A Euclidean distance between pair of clusters can be represented by a corresponding weighted edge. Our Algorithm is also based on the minimum spanning tree but not limited to two-dimensional points. There were two kinds of clustering problem; one that minimizes the maximum intra-cluster distance and the other maximizes the minimum inter-cluster distances. Our Algorithm produces clusters with both intra-cluster and inter-cluster similarity. The Second phase of the algorithm converts the minimum spanning tree **EMST2** into dendrogram, which can be used to

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

interpret about inter-cluster distances. This new algorithm is neither single link clustering algorithm (SLCA) nor complete link clustering algorithm (CLCA) type of hierarchical clustering, but it is based on the distance between centers of clusters. This approach leads to new developments in hierarchical clustering. The level function, *L*, records the proximity at which each clustering is formed. The levels in the dendrogram tell us the least amount of similarity that points between clusters differ. This piece of information can be very useful in several medical and image processing applications.

Here, we use a cluster validation criterion based on the geometric characteristics of the clusters, in which only the inter-cluster metric is used. The **DGMST** algorithm is a nearest centroid-based clustering algorithm, which creates region or subtrees (clusters/regions) of the data space. The algorithm partitions a set *S* of data of data *D* in data space in to *n* regions (clusters). Each region is represented by a centroid reference vector. If we let *p* be the centroid representing a region (cluster), all data within the region (cluster) are closer to the centroid *p* of the region than to any other centroid *q*:

$$R(p) = \{x \in D \;/\; dist(x, p) \le dist(x, q) \;\; \forall q\}$$

Thus, the problem of finding the proper number of clusters of a dataset can be transformed into problem of finding the proper region (clusters) of the dataset. Here, we use the **MST** as a criterion to test the inter-cluster property. Based on this observation, we use a cluster validation criterion, called Cluster Separation (CS) in **DGMST** algorithm [4].

Cluster separation (CS) is defined as the ratio between minimum and maximum edge of MST. ie

$$CS = E_{min} / E_{max},$$

where E_{max} is the maximum length edge of **MST**, which represents two centroids that are at maximum separation, and E_{min} is the minimum length edge in the MST, which represents two centroids that are nearest to each other. Then, the CS represents the relative separation of centroids. The value of CS ranges from 0 to 1. A low value of CS means that the two centroids are too close to each other and the corresponding partition is not valid. A high CS value means the partitions of the data is even and valid. In practice, we predefine a threshold to test the CS. If the CS is greater than the threshold, the partition of the dataset is valid. Then again partitions the data set by creating subtree (cluster/region). This process continues until the CS is smaller than the threshold. At that point, the proper number of clusters will be the number of cluster minus one. The CS criterion finds the proper binary relationship among clusters in the data space. The value setting of the threshold for the CS will be practical and is dependent on the dataset. The higher the value of the threshold the smaller the number of clusters would be. Generally, the value of the threshold will be > 0.8[4]. Figure 3 shows the CS value versus the number of clusters in hierarchical clustering. The CS value < 0.8 when the number of clusters is 5. Thus, the proper number of clusters for the data set is 4. Further more, the computational cost of CS is much lighter because the number of subclusters is small. This makes the CS criterion practical for the **DGMST** algorithm when it is used for clustering large dataset.

Algorithm: DGMST ( )
 Input : *S* the point set
 Output : dendrogram with optimal number of clusters
Let *e1* be an edge in the **EMST1** constructed from *S*
Let *e2* be an edge in the **EMST2** constructed from C
Let W_e be the weight of *e1*
Let σ be the standard deviation of the edge weights in **EMST1**
Let S_T be the set of disjoint subtrees of **EMST1**
Let n_c be the number of clusters

1. Construct an **EMST1** from *S*
2. Compute the average weight of $\hat{W}$ of all the Edges from **EMST1**
3. Compute standard deviation σ of the edges from **EMST1**
4. $S_T = ø$; $n_c = 1$; $C = ø$;
5. **Repeat**
6. **For** each $e1 \in$ **EMST1**
7. **If** $(W_e > \hat{W} + \sigma)$ or (current longest edge *e1)*
8. Remove *e1* from **EMST1**
9. $S_T = S_T \cup \{ T' \}$ // *T*'' is new disjoint Subtree (regions)
10. $n_c = n_c+1$
11. Compute the center C_i of T_i using eccentricity of points
12. $C = \cup_{T_i \in S_T} \{C_i\}$
13. Construct an **EMST2** *T* from *C*
14. E_{min} = get-min-edge (*T*)
15. E_{max} = get-max-edge (*T*)

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

16. CS = E_{min} / E_{max}
17. **Until** CS < 0.8
18. Begin with disjoint clusters with level *L* (0) = 0 and sequence number *m* = 0
19. **While** (*T* has some edge)
20. *e2* = get-min-edge(*T*) // for least dissimilar pair of clusters
21. (*i*, *j*) = get-vertices (*e2*)
22. Increment the sequence number *m* = *m* + 1, merge the clusters (*i*) and (*j*), into single cluster to form next clustering *m* and set the level of this cluster to L(*m*) = *e2*;
23. Update *T* by forming new vertex by combining the vertices *i*, *j*;
24. **Return** dendrogram with optimal number of clusters

Figure 1 shows a typical example of **EMST1** constructed from point set S, in which inconsistent edges are removed to create subtree (clusters/regions). Our algorithm finds the center of the each cluster, which will be useful in many applications. Our algorithm will find optimal number of clusters or cluster structures. Figure 2 shows the possible distribution of the points in the two cluster structures with their center vertex as 5 and 3.

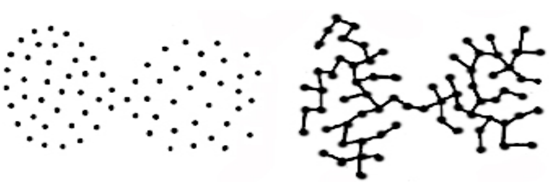
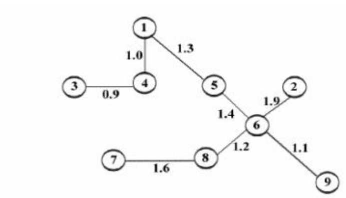
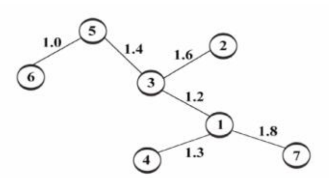


Figure 1: Clusters connected through points -EMST1



Vertex	1	2	3	4	5	6	7	8	9
Eccentricity	5.5	6.7	7.4	6.5	4.2	4.6	7.4	5.8	5.7



Vertex	1	2	3	4	5	6	7
Eccentricity	3.6	4.6	3.0	4.9	4.4	5.4	5.4

Figure 2: Two Clusters/regions with Center points 5 and 3

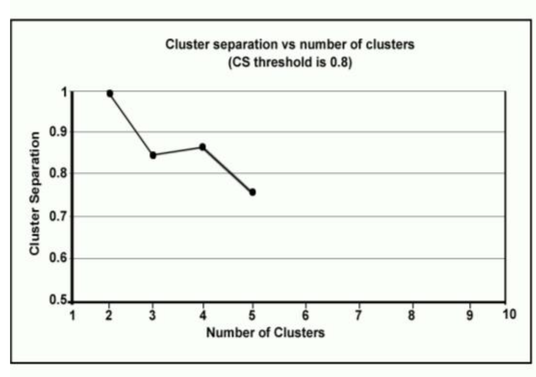


Figure 3: Number of Clusters vs. Cluster Separation

Our **DGMST** algorithm works in two phases. The outcome of the first phase (lines 1-17) of the algorithm consists of optimal number clusters with their center. It first constructs **EMST1** form set of point S (line 1). Average weight of edges and standard deviation are computed (lines 2-3). Inconsistent edges are identified and removed from **EMST1** to generate subtree *T'* (lines 7-9). The center for each subtree (cluster/region) is computed at line 11. Using the cluster/region center point again another minimum spanning tree **EMST2** is constructed (line 13). Using the new evaluation criteria, optimal number of clusters/regions is identified (lines 14-16). Lines 6-16 in the algorithm are repeated until optimal number of clusters are obtained. We use the graph of Figure 4 as example to illustrate the second phase (lines 18-24) of the algorithm.

The second phase of the **DGMST** algorithm construct minimum spanning tree *T* from the point set $C = \{c_1, c_2, c_3 \ldots c_k\}$ and convert the *T* into dendrogram is shown in figure 5. It places the entire disjoint cluster at level 0 (line 18). It then checks to see if *T* still contains some edge (line 19). If so, it finds minimum edge *e2* (line 20). It then finds the vertices i, j of an edge *e2* (line 21). It then merges the vertices (clusters) and forms a new vertex (*agglomerative approach*). At the same time the sequence number is increased by one and the level of the new cluster is set to the edge weight (line 22). Finally, the Updation of minimum spanning tree is performed at line 23. The lines 20-23 in the algorithm are repeated until the minimum spanning tree *T* has no edge. The dendrogram with optimal number of cluster as objects is generated. The objects within the clusters are compact. The clusters are well separated, shown in Figure 4.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

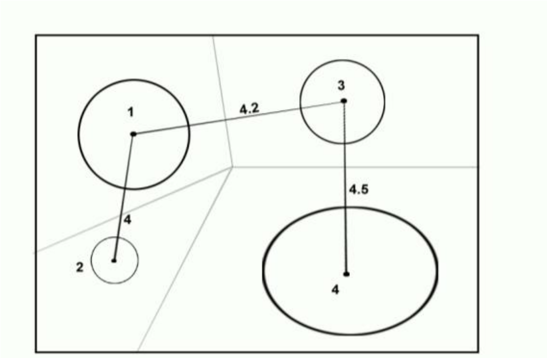


Figure 4. EMST2 From 4 region/cluster center points

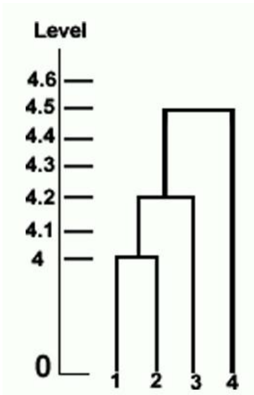


Figure 5. Dendrogram for Optimal Meta cluster

IV. CONCLUSION

Our **DGMST** clustering algorithm does not assumes any predefined cluster number. The algorithm gradually finds clusters with center for each cluster. These clusters ensure guaranteed intra-cluster similarity. Our algorithm does not require the users to select and try various parameters combinations in order to get the desired output. Our **DGMST** clustering algorithm uses a new cluster validation criterion based on the geometric property of partitioned regions/clusters to produce optimal number of "true" clusters with center for each of them. Our algorithm also generates *dendrogram* which is used to find the relationship between the optimal number clusters. The inter-cluster distances between clusters/regions are shown in the Dendrogram. This will be very useful in many applications. All of these look nice from theoretical point of view. However from practical point of view, there is still some room for improvement for running time of the clustering algorithm. This could perhaps be accomplished by using some appropriate data structure. In the future we will explore and test our proposed clustering algorithm in various domains. The **DGMST** algorithm uses both divisive and agglomerative approach to find *Optimal Meta similarity clusters*. We will further study the rich properties of EMST-based clustering methods in solving different clustering problems.

REFERENCES

[1] Anil K. Jain, Richard C. Dubes "Algorithm for Clustering Data", *Michigan State University, Prentice Hall, Englewood Cliffs, New Jersey* 07632.1988.

[2] T. Asano, B. Bhattacharya, M.Keil and F.Yao. "Clustering Algorithms based on minimum and maximum spanning trees". *In Proceedings of the 4th Annual Symposium on Computational Geometry*,Pages 252-257, 1988.

[3] D. Avis "Diameter partitioning." *Discrete and Computational Geometr*, 1:265-276, 1986.

[4] Feng Luo,Latifur Kahn, Farokh Bastani, I-Ling Yen, and Jizhong Zhou, "A dynamically growing self-organizing tree(DGOST) for hierarchical gene expression profile",Bioinformatics,Vol 20,no 16, pp 2605-2617, 2004.

[5] M. Fredman and D. Willard. "Trans-dichotomous algorithms for minimum spanning trees and shortest paths". In *Proceedings of the 31st Annual IEEE Symposium on Foundations of Computer Science*,pages 719-725, 1990.

[6] M. Halkidi, Y.Batistakis and M. Vazirgiannis "On clustering validation techniques", *J.Intel. Inform. System.*, 17, 107-145, 2001

[7] M. Halkidi, Y.Batistakis and M. Vazirgiannis, "Clustering validity checking methods:part II" *SIGMOD record.*, 31, 19-27, 2002

[8] G. Hamerly and C. Elkan, "Learning the k in k-means, in Advances in Neural Information Processing Systems" 16, S. Thrun, L. Saul, and B. Schölkopf, eds., *MIT Press, Cambridge, MA*, 2004.

[9] P. Hansen and M. Delattre, "Complete-link cluster analysis by graph coloring" *Journal of the American Statistical Association* 73, 397-403, 1978.

[10] A. Hardy, "On the number of clusters", *Computational Statistics and Data Analysis*, 23, pp. 83–96, 1996.

[11] T. Hastie, R. Tibshirani, and J. Friedman, "The elements of statistical learning: Data mining, inference and prediction", *Springer-Verlag, 2001*.

[12] Hubert L. J "Min and max hierarchical clustering using asymmetric similarity measures" *Psychometrika* 38, 63-72, 1973.

[13] H.Gabow, T.Spencer and R.Rarjan. "Efficient algorithms for finding minimum spanning trees in undirected and directed graphs", *Combinatorica,* 6(2):109-122, 1986.

[14] J.C. Gower and G.J.S. Ross "Minimum Spanning trees and single-linkage cluster analysis" *Applied Statistics* 18, 54-64, 1969.

[15] S. John Peter, S.P. Victor, "A Novel Algorithm for Meta similarity clusters using Minimum spanning tree". *International Journal of computer science and Network Security*. Vol.10 No.2 pp. 254 – 259, 2010

[16] S. C. Johnson, "Hierarchical clustering schemes" *Psychometrika* 32, 241-254, 1967.

[17] D. Johnson, "The np-completeness column: An ongoing guide". *Journal of Algorithms*,3:182-195, 1982.

[18] D. Karger, P. Klein and R. Tarjan, "A randomized linear-time algorithm to find minimum spanning trees". *Journal of the ACM*, 42(2):321-328, 1995.

[19] J. Kruskal, "On the shortest spanning subtree and the travelling salesman problem", *In Proceedings of the American Mathematical Society,* pp 48-50, 1956.

[20] J. Nesetril, E.Milkova and H.Nesetrilova. Otakar boruvka on "Minimum spanning tree problem": Translation of both the 1926 papers, comments, history. DMATH*: Discrete Mathematics,* 233, 2001.

[21] Oleksandr Grygorash, Yan Zhou, Zach Jorgensen. "Minimum spanning Tree Based Clustering Algorithms". *Proceedings of the 18th IEEE International conference on tools with Artificial Intelligence (ICTAI'06)* 2006.

[22] N. Paivinen, "Clustering with a minimum spanning of scale-free-like structure".*Pattern Recogn. Lett.*,26(7): 921-930, 2005.

[23] F. Preparata and M.Shamos. "Computational Geometry": An Introduction. *Springer-Verlag, Newyr, NY ,USA*, 1985

[24] R. Prim. "Shortest connection networks and some generalization". *Bell systems Technical Journal*,36:1389-1401, 1957.

[25] R. Rezaee, B.P.F. Lelie and J.H.C. Reiber, "A new cluster validity index for the fuzzy C-mean", *Pattern Recog. Lett., 19*,237-246, 1998.

[26] D. M. Rocke and J. J. Dai, "Sampling and subsampling for cluster analysis in data mining: With applications to sky survey data", *Data Mining and Knowledge Discovery,* 7, pp. 215–232, 2003.

[27] S. Salvador and P. Chan, "Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms", *in Proceedings Sixteenth IEEE International Conference on Tools with Artificial Intelligence, ICTAI 2004, Los Alamitos, CA, USA, IEEE Computer Society*, pp. 576–584 , 2004.

[28] S. Still and W. Bialek, "How many clusters?*" , An information-theoretic perspective*, *Neural Computation*, 16, pp. 2483–2506, 2004.

[29] Stefan Wuchty and Peter F. Stadler. "Centers of Complex Networks". 2006

[30] C. Sugar and G. James, "Finding the number of clusters in a data set *", An information theoretic approach*, Journal of the American Statistical Association, 98 pp. 750–763, 2003.

[31] S. Theodoridis, K. Koutroubas, "Pattern recognition" *Academic Press,* 1999

[32] R. Tibshirani, G. Walther and T.Hastie "Estimating the number of clusters in a dataset via the gap statistic". *J.R. Stat. Soc.Ser.B,63*.411-423, 2001.

[33] Y.Xu, V.Olman and D.Xu. "Minimum spanning trees for gene expression data clustering". *Genome Informatics*,12:24-33, 2001.

[34] C. Zahn. "Graph-theoretical methods for detecting and describing gestalt clusters". *IEEE Transactions on Computers*, C-20:68-86, 1971.

BIOGRAPHY OF AUTHORS

S. John Peter is working as Assistant professor in Computer Science, St.Xavier's college (Autonomous), Palayamkottai, Tirunelveli. He earned his M.Sc degree from Bharadhidasan University, Trichirappalli. He also earned his M.Phil from Bharadhidasan University, Trichirappalli. Now he is doing Ph.D in Computer Science at Manonmaniam Sundranar University, Tirunelveli, Tamil Nadu, - INDIA. He has published research papers on clustering algorithm in various national and international Journals.

E-mail: jaypeeyes@rediffmail.com

Dr. S. P. Victor earned his M.C.A. degree from Bharathidasan University, Tiruchirappalli. The M. S. University, Tirunelveli, awarded him Ph.D. degree in Computer Science for his research in Parallel Algorithms. He is the Head of the department of computer science, and the Director of the computer science research centre, St. Xavier's college (Autonomous), Palayamkottai, Tirunelveli. The M.S. University, Tirunelveli and Bharathiar University, Coimbatore have recognized him as a research guide. He has published research papers in international, national journals and conference proceedings. He has organized Conferences and Seminars at national and state level.
E-mail: victorsp@rediffmail.com

 http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Media Streaming using Multiple Description Coding in Overlay Networks

Sachin Yadav
Department of CSE
SGIT College of Engineering
Ghaziabad, India
sac.yaduvanshi@gmail.com

Ranjeeta Yadav
Department of ECE
SGIT College of Engineering
Ghaziabad, India
ranjeeta29@gmail.com

Shailendra Mishra
Department of CSE
Kumaon Engineering College
Dwarahat, India
skmishra1@gmail.com

Abstract—**In this paper we examine the performance of two types of Overlay networks i.e. Peer-to-Peer (P2P) & Content Delivery Network (CDN) media streaming using Multiple Description Coding (MDC). In both the approaches many servers simultaneously serve one requesting client with complementary descriptions. This approach improves reliability and decreases the data rate a server has to provide. We have implemented both approaches in the ns-2 network simulator. The experimental results indicate that the performance of Multiple Description Coding-based media streaming in case of P2P network is better than CDN.**

Keywords- MDC; CDN; Video Streaming; P2P; Overlay Network

I. INTRODUCTION

Media streaming received lot of attention in the past few years. As a consequence, live and on-demand media streaming is today widely used to stream TV & radio channels, TV shows, or arbitrary audio & video media. During this time several approaches have been devised to tackle the media-streaming problem. The first one is to use a client-server model, where a single server is the media provider and multiple clients are the media consumers. The second one is to use a peer-to-peer approach where the clients help the server in delivering the media content by having the roles of consumers and providers at the same time.

Both schemes have their advantages and disadvantages. The client-server approach has the advantage that the client receives the content directly from the server with the minimum delay but at the cost of overwhelming the server in particular situations (for instance at high rate hours: e.g. football / basketball games etc). As a result, the server's bandwidth can quickly become a bottleneck in the system due to the large number of client requests. On the other hand, in the peer-to-peer approach algorithms are devised to multicast the content between clients. In this case the clients have an active role in distributing the media content to other clients and thus remove the pressure from the server node. In this way, scaling the system functionality to a large number of consumers becomes a reality. However, this solution has its drawbacks too. Specifically, these algorithms have to tackle a high dynamic system, where clients can come and leave suddenly without any prior knowledge or guarantees.

Today's video streaming systems are mostly based on the client server model of Content Delivery Networks (CDN) which leads to several problems. The most important ones are:

1. *Flash Crowd:* Large numbers of streaming servers are not able to feed more than a few hundred streaming sessions simultaneously [2].
2. *Bandwidth cost:* It can be a significant problem to the content provider. In contrast, these costs are shared by every participant in the P2P streaming network.
3. *Single Point of failure:* Like any client-server model, the server is the single point of failure.

P2P networks offer characteristics and possibilities which cannot be provided by CDNs as proposed in [10]. As we show in this work, the performance of media streaming can he better in a P2P network, although the probability that one stream breaks is higher [13] [4]. The reason for this is that the replication rate of the video streams in a P2P network is typically significantly higher than in a CDN, due to the large number of participating hosts. In Gnutella for example, every peer shares an average of 500 files [15] and many peers host the same file.

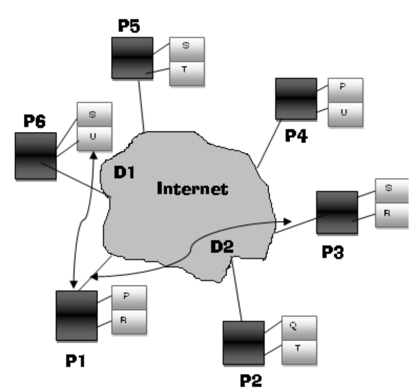


Figure 1: Distributed video streaming using multiple description coding in a P2P network. Peer P1 is

simultaneously serve by the closest available peers P6 & P3 with descriptions D1 & D2 respectively.

Using MDC in a P2P streaming scenario is illustrated in Figure 1. Peer p1 wants to receive video file S which is available in the MDC format on p3, p5 and p6. In this example the video is encoded using two descriptions D1 & D2. Peers p3 and p6 are chosen based on the distance from server to the receiver, and they simultaneously serve the video file S, each one providing a complementary description. If both the descriptions are received at the receiving peer p1, it will experience the highest quality. If any of the descriptions are affected by packet loss or excessive delay, the receiver can still decode and display video S but at the expense of a degradation of the quality, as the descriptions are independently decodable.

II. Multiple Description Video Coding

Multiple Description coding (MDC) is a coding technique that fragments a single media stream into n sub streams ($n \geq 2$) referred to as descriptions. The packets of each description are routed over multiple, (partially) disjoint paths. In order to decode the media stream, any description can be used, however, the quality improves with the number of descriptions received in parallel. The idea of MDC is to provide error resilience to media streams. Since an arbitrary subset of descriptions can be used to decode the original stream, network congestion or packet loss — which are common in best-effort networks such as the Internet — will not interrupt the stream but only cause a (temporary) loss of quality. The quality of a stream can be expected to be roughly proportional to data rate sustained by the receiver.

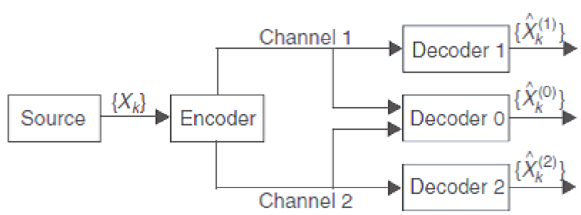


Figure 2: MD source coding with two channels and three receivers. The general case has M channels and 2^{M-1} receivers.

This property makes MDC highly suitable for lossy packet networks where there is no prioritization among the packets. The principle of MDC encoding/decoding is illustrated in figure 2. For a general overview on Multiple Description Coding (MDC) refer to [1].

III. Video Streaming Over Internet

Media streaming systems are distinct from the file sharing systems [11], in which a client has to download the entire file before using it. Real-time multimedia, as the name implies, has timing constraints. For example, audio and video data must be played out continuously. If the data does not arrive in time, the play out process will pause, which is annoying to human ears and eyes. Real-time transport of live video or stored video is the predominant part of real-time multimedia. In this paper, we are concerned with video streaming, which refers to real-time transmission of stored video. There are two modes for transmission of stored video over the Internet, namely the download mode and the streaming mode (i.e., video streaming). In streaming mode, the video content need not be downloaded in full, but is being played out while parts of the content are being received and decoded. Due to its real-time nature, video streaming typically has bandwidth, delay and loss requirements. However, the current best-effort Internet does not offer any quality of service (QoS) guarantees to streaming video over the Internet. In addition, for multicast, it is difficult to efficiently support multicast video while providing service flexibility to meet a wide range of QoS requirements from the users. Thus, designing mechanisms and protocols for Internet streaming video poses many challenges. It has been demonstrated in [16] that using MDC in combination with packet path diversity significantly improves the robustness of a real-time video application.

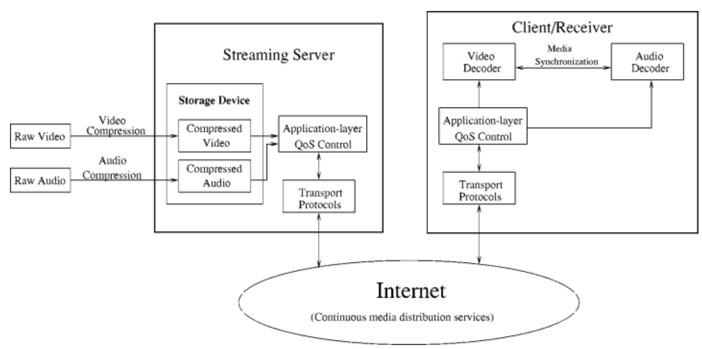


Figure 3: Architecture for video streaming.

In Figure 3, raw video and audio data are pre-compressed by video compression and audio compression algorithms and then saved in storage devices. Upon the client's request, a streaming server retrieves compressed video/audio data from storage devices and then the application-layer QoS control module adapts the video/audio bit-streams according to the network status and QoS requirements. After the adaptation, the transport protocols packetize the compressed bit-streams and send the video/audio packets to the Internet. Packets may be dropped or experience excessive delay inside the Internet due to congestion. To improve the quality of video/audio transmission, continuous media distribution services (e.g., caching) are deployed in the Internet. For packets that are successfully delivered to the receiver, they first pass through the transport layers and are then processed by the application layer before being decoded at the video/audio decoder. To achieve synchronization between video and audio presentations, media synchronization mechanisms are required. From Figure 3, it can be seen that the six areas are closely related and they are coherent constituents of the video streaming architecture.

IV. Modelling

We use the following methodologies in our simulations to reflect the real-world network situations.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

1. Modeling Availability in P2P Networks

In P2P networks, peer and content availability poses a challenging problem to be solved. Availability of a peer in a P2P network is quite unpredictable, depending primarily on human presence. In our experiments we model peer availability as a 2 state markov process, having the states ON and OFF. The average lifetime of a peer in a Gnutella network is found to be about 30 minutes [14]. For our experiments we take a Gaussian distribution of ON time, which has a mean of 30 minutes. To model the availability of content among the peers, we randomly choose peers having a particular media file. We vary the percentage of peers having the file from 5% to 50%.

2. Server Placement in CDN

The server placement problem addresses how to optimally place a number of servers in order to maximize the quality at the end user. In our experiments we varied the number of servers to obtain measurement of Quality of Service, such as packet loss and response time. For a particular number of servers, we placed the servers randomly in the network and measured the average round-trip-time from each user to the servers. We performed this random placement 10 times and chose the one yielding the smallest average round-trip-time.

3. Server Selection in P2P and CDN Network

The server selection problem addresses how to optimally choose a pair of servers to get complementary descriptions in order to maximize the perceived quality at the receiver. As described in [10] Apostolopoulos proposed a path diversity model which requires the knowledge of network topology, including knowledge of joint and disjoint links, and loss characteristics for each link. In our experiments we simply choose the closest two servers for each client request. For P2P case, we choose the closest two serving peers having the required content.

4. Content Distribution across Servers in CDN

This problem addresses how to optimally distribute the Multiple Description streams in an existing set of servers. In this paper we assume that all the CDN servers contain both the descriptions, which simplifies the server selection problem by merely choosing the two closest servers.

5. Network Load

To simulate the network load, we created random TCP connections originating from arbitrary nodes, on the average 3 new connections per second, each connection lasting for 1 minute.

V. RESULTS

We implemented both the P2P and CDN approaches within the Network Simulator ns-2 [8]. The topology was created using the GT-ITM topology generation tool with the transit-stub model, having 100 nodes. A video file of 1 minute duration, having a data-rate of 100 Kbit/s was selected for all the simulations. Each packet contains 1000 bytes. In both the CDN and P2P based systems, there is one new request every second, originating from an arbitrary node. In P2P network, the file is streamed from two closest available peer nodes with complementary descriptions, whereas in CDN, the same is served by two closest CDN servers. It was assumed that a peer can serve only one request at one time, while a CDN server can serve a maximum of 200 streams simultaneously.

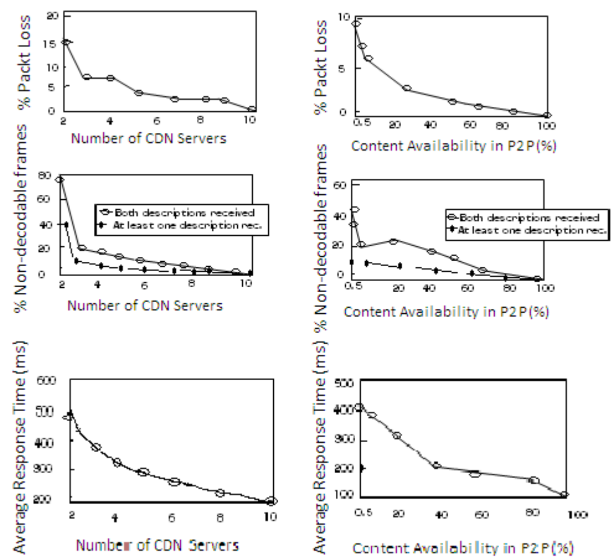


Figure 4: Performance of P2P and CDN networks using MDC: (top) packet loss rate varies with varying number of CDN servers and content availability in p2p network. (middle) number of decodable frames increases with increasing number of servers and availability. (bottom) average response times for P2P and CDN.

Figure 4 shows the results obtained through simulations. Three performance parameters, namely the rate of packet loss, number of non-decodable frames and the average response time, i.e. the time to receive the first video packet after the request has been sent, are compared for P2P and CDN networks. For the count of non-decodable frames, it is assumed that the descriptions contain an Intra frame once in every second, and in case of a packet loss for the P-frames, all the subsequent frames become non-decodable, until the next I-frame is received. Because of MDC coding, the receiver can still view with a reduced frame rate, unless both the descriptions are corrupted simultaneously. This is shown in figure 5, where description s1 contains a packet loss, but s2 is received error-free. The receiver can view with ½ the original frame rate until the next I-frame is received in s1.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

The simulation results indicate that the performance of a P2P network is comparable to that of a CDN, even at the high unavailability of peers and content in the p2p network.

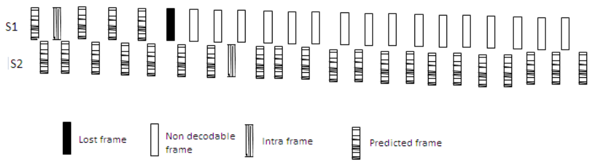


Figure 5: Impact of packet loss in MDC-based video streaming. Only stream s2 can be decoded completely. s1 is affected by packet loss and lead to locally reduced frame rate of the reconstructed video.

VI. Related Work

Peer-to-peer based media streaming approaches using multiple serving hosts have been proposed in [12] and [5]. In [16] MDC-based distributed video streaming has been proposed for content delivery networks. Our work is inspired by this work and we use the same multiple description encoding technique for a P2P network.

VII. Conclusion

In this paper we presented a performance comparison of P2P media streaming with CDN – based media streaming, both employing MDC. The P2P approach takes advantage of multiple supplying peers to combat the inherent limitations of the P2P network and the best effort Internet. The media content is encoded using a multiple description encoder which allows realizing distributed streaming from more than one peer. In the final paper we plan to also provide experimental results on video dispersion, i.e. the time it takes to be able to satisfy a large number of streaming requests for a new video that is injected into the network, for both the P2P and CDN network.

References

[1] V. K. Goyal. “Multiple Description Coding: Compression Meets the Network”, IEEE Signal Processing Magazine, vol. 18, n.5, page(s) 74 – 94. September 2001.

[2] Tyron Stading, Petros Maniatis, Mary Baker, “Peer-to-Peer Caching Schemes to Address Flash Crowds”, 1st International Peer To Peer Systems Workshop (IPTPS) 2002.

[3] V. Padmanabhan, H. Wang, P. Chou, and K. Sripanidkulchai. “Distributing streaming media content using cooperative networking.” In Proc of NOSSDAV’02, Miami Beach, FL, USA, May 2002.

[4] S. Saroiu, P. K. Gummadi, and S. D. Gribble, “A measurement study of peer-to-peer file sharing systems”. In Proceedings of Multimedia Computing and Networking (MMCN), Jan. 2002.

[5] T. Nguyen and A. Zakhor, “Distributed VideoStreaming”, Multimedia Computing and Networking 2002, Multimedia Conference, Jan 2002.

[6] Borella, M. S., D. Swider, S. Uludag, and G. B. Brewster, “Internet Packet Loss: Measurement and Implications for End-to- End QoS”, in Proceedings of the International Conference on Parallel Processing (1998).

[7] F. M. Cuenca-Acuna, R. P. Martin, T. D. Nguyen. “Autonomous Replication for High Availability in Unstructured P2P Systems”, Technical Report DCS-TR-509, Department of Computer Science, Rutgers University. April 2003.

[8] The Network Simulator, http://www.isi.edu/nsnam/ns/

[9] J. Kangasharju, K.W. Ross, D. Turner, “Adaptive Replication and Replacement in P2P Caches”, 2002. working paper.

[10] J. Apostolopoulos, T. Wong, W. Tan, S. Wee, “On Multiple Description Streaming with Content Delivery Networks”, IEEE INFOCOM, June 2002.

[11] D. Xu, M. Hefeeda, S. Hambrusch, and B. Bhargava. “On peer-to-peer media streaming.” In Proc. of IEEE ICDCS, Vienna, Austria, July 2002.

[12] M. Hefeeda, B. Bhargava “On-Demand Media Streaming Over the Internet”, CERIAS TR 2002- 20, Purdue University, June 2002.

[13] R. Baghwan, S. Savage, G, M. Voelker. “Understanding Availability”. In Proceedings of the International Workshop on Peer-to-Peer Systems (IPTPS). February 2003.

[14] E. Adar, B. Huberman. “Free Riding on Gnutella”, First Monday, 5(10). October 2000.

[15] J. Chu, K. Labonte, B. N. Levine, “Availability and locality measurements of peer-to-peer file systems,” in ITCom: Scalability and Traffic Control in IP Networks, Proceedings of SPIE. vol. 4868, Jul. 2002.

[16] J. G. Apostolopoulos, “Reliable video communication over lossy packet networks using multiple state encoding and path diversity”, VCIP, January 2001.

Authors Profile

Sachin Yadav received his Bachelors in Computer Engineering from Pune University in 2001. He is working currently as a Associate Prof & Head for the computer science department in SGIT, Ghaziabad and is pursuing his PhD from the Uttarakhand Technical University, Dehradun.

Ranjeeta Yadav received her Bachelors in Electronics & Communication Engineering from Uttar Pradesh Technical University in 2004. She received her M.Tech in signal processing from Delhi University in 2009. She is working currently as an Assistant Prof in Electronics & Communication Engineering department in SGIT, Ghaziabad.

Shailendra Mishra got his Master degree from MNNIT, Allahabad and Ph.D in Computer Science & Engineering from Gurukul Kangri University, Haridwar. He is working currently as a Prof & Head for the Computer Science department in KEC, Dwarahat. He published several research papers in international journals and international conferences.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Secured and QoS based multicast routing in MANETs

Maya Mohan
Department of CSE
NSS College of Engineering
Palakkad, Kerala.
mayajeevan@gmail.com

S.Mary Saira Bhanu
Department of CSE
National Institute of Technology
Thiruchirappalli, TN.
msb@nitt.edu

***Abstract*- A mobile ad-hoc network (MANET) is a dynamic network of self controlled mobile nodes without any centralized co-ordinator (access point or base station) or wired infrastructure. The main difficulty in designing a routing protocol for MANETs is the dynamical topology which results from the random movement of mobile nodes within the source's transmission range. MANET, which is fundamentally different from conventional infrastructure based networks, is self-configuring and formed directly by a set of mobile nodes. In MANET, the heterogeneity of networks and destinations makes it difficult to improve bandwidth utilization and service flexibility. Therefore, mobility of nodes makes the design of data distribution jobs greatly challenging. The wide use of multiparty conferences in MANETs leads to multicast routing for the transmission of information, such as video and other streaming data. In multicasting quality of service (QoS) and security are the leading challenges. The QoS deals with bandwidth utilization and network failures and security provides group communication to be confidential. In this paper MAODV protocol is modified by including QoS as well as security to the group communication. The QoS includes the link failures and the node failures. The security is provided by using symmetric key encryption method.**

Key Words- multicast; MANET; QoS; security;

I. INTRODUCTION

A mobile ad-hoc network (MANET) is a collection of wireless mobile nodes that forms a dynamic network without any centralized coordinator. The highlighted application areas of MANETs are rescue sites, battlefields, group conferences etc. Communication in MANETs is provided by a sequence of neighbor nodes from a source to a destination form a path and intermediate mobile nodes relay packets in a store-and–forward mode. Some typical applications of MANETs, nodes need to accomplish a task by group. Therefore, the multicasting plays a significant role in the MANETs. Multicasting reduces the communication cost for applications that sends the same data to many recipients. It reduces the channel bandwidth, processing time between sender and router and delivery delay by sending the data simultaneously to different recipients instead of multiple unicasts. In addition, it gives robust communication whereby the receiver address is unknown or modified without the knowledge of the source within the wireless environment [1].

Network researches have been done in the area of quality of service (QoS) and security with few exceptions. However, security [25] impacts the overall network QoS as more security usually means more message overhead for authentication and other security functions, as well as additional delays imposed due to overhead caused by encryption, etc. This is especially true in an ad- hoc network environment where security mechanisms [5] such as authentication services are proposed to protect the communication on open medium in wireless networks, thus introducing overhead that affect the QoS of communications significantly. It is therefore essential to consider both security and QoS together when designing protocols for ad- hoc environments as one impact the other. Very little work has been done on the interaction between security and QoS in wireless networks.

In this paper an effort has been taken in order to provide security as well as quality of service for group communication. QoS [4] includes handling node failures, link failures and finding the path when node mobility occurs and storing the data while unavailability of paths. It also saves the bandwidth by using less control messages by including symmetric key encryption while compared with asymmetric key encryption

.In ad- hoc networks, users need to assure the party who supposedly sent a message to another party is indeed the legitimate party. Otherwise, a malicious node could tamper a network with falsified data. These attacks can result in degraded performance of networks, interference of resource

reservation, and unauthorized use of resources. There are two basic kinds of cryptography that have been widely used for the networks: *symmetric* cryptography and *asymmetric* cryptography (such as digital signature).

The communication links in MANETs are open shared medium, which makes the communications between neighboring nodes more vulnerable to attacks such as packet forging and malicious alteration. In addition, MANETs are characterized by absence of fixed infrastructure, rapid topology change and constrained resources (such as limited battery power, small computational capacity and bandwidth). These characteristics determine that the authentication protocols used for routing and data packet delivery in MANETs should be lightweight and scalable. Asymmetric cryptography does not adapt well to MANETs in that the processing required for asymmetric cryptography is very high and the technique has been proved to be prohibitively insufficient in wireless ad- hoc networks in terms of message overhead and computation complexity. Symmetric cryptography algorithms are fast. Even though it introduces complexity in key maintenance but needs less computational power which in turn saves life of battery.

The basic principle of Multicast Ad-hoc on demand distance vector (MAODV) [2] is adopted from AODV [3]. The security and QoS features are added to MAODV.. The MAODV protocol is opted because of its medium node and network overhead. The routing table of MAODV contains only the next hop address not the entire route which helps in saving the cache memory. Periodic updates are not happening in MAODV which will help to reduce the control messages.

The rest of the paper is structured as follows: Section II discusses the previous efforts in this area. Section III dictates the operational principle of MAODV, the security measures taken for secure group communication and the QoS measures taken to incorporate security and section IV deals with the simulation results. Section V concludes the work.

II. RELATED WORK

Multicasting plays a critical role in group conferences, multiparty games etc. A comparative study is carried out with different multicast routing protocols in ad-hoc networks [16]. A performance comparison of MAODV and ODMRP is explained in [18].

The specific security requirements of MANETs (in particular, key management) are still considered to be open research challenges. Recently, several key agreement protocols for MANETs were proposed [6]. Mobility impacts performance only when members cross groups. For instance, when two partners provide broadcast services for users in two overlapping groups, users moving within each group are managed by their local group key distributors (GKDs) and without any coordination between their broadcasts. On the other hand, when a user crosses from one group to another, security should be transferred between partners. A comparative study has been done based on different security mechanisms in MANETs [17]. One of the security mechanisms in multicasting is group re-keying [19], which is an efficient and scalable mechanism that exploits the property of ad hoc networks in which each member of a group is both a host and a router, and distributes the group key to member nodes via a secure hop-by-hop propagation scheme. A probabilistic scheme based on pre-deployed symmetric keys is used for implementing secure channels between members for group key distribution. In MANETs, the computational load and complexity for key management is strongly subjected to restriction of the node's available resource and the dynamic nature of network topology. Secure and Efficient Key Management (SEKM) [19] is an efficient method for MANETs. In SEKM, the server group creates a view of the CA (Certifying Authority) and provides certificate update service for all nodes, including the servers themselves. A ticket scheme is introduced for efficient certificate service. In SEKM, server group is formed securely and maintains connectivity. The certificate-updating request is processed by server group in a ticket-based approach. The system secret, held by each server, is refreshed periodically in a fair and efficient easy. The public key mechanism used above increases the computational complexity.

Apart from security, wide range of work has been done in the area of QoS. The problem of QoS routing in wired networks is not similar in a dynamic network environment [21], especially the application of these algorithms in a MANET. QoS-AODV [22] has been proposed for QoS extension requirement, but it does not consider the best route. Indeed, it chooses the minimum delay and hop count route. SQoS [23] is a secure form of QoS-Guided Route Discovery for on-demand ad hoc network routing. In [24], a flexible QoS model for MANETs (FQMM) is explained, which is a hybrid service model and based on IntServ and Diffserv model. FQMM combines the reservation procedure for high priority traffic with service differentiation for low-priority traffic. Thus, FQMM provides the ideal QoS for per flow and overcomes the scalability problem by classifying the low-priority traffic into service classes. Less security measures are adopted in FQMM. QoS mainly deals with end to end delay and bandwidth. QoS provides a set of service requirements to the flows while routing them through the network [7]. The widespread use of wireless technologies has increased QoS for multimedia applications in wireless networks and traditional internet QoS protocols like RSVP [8] cannot be used for wireless environment due to the error-prone nature of wireless links and the high mobility of mobile devices in MANETs. Therefore, providing QoS in MANETs is more challenging than in fixed and wireless networks. In order to overcome the above drawbacks a new proposal for QoS and security based on MAODV is introduced. The protocol identifies node failures and link failures which is not covered by the above

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

entioned mechanisms and also provides security for the data transmitting.

III. OPERATIONAL PRINCIPLES

A. MAODV

MAODV is the multicast extension of AODV. Both AODV and MAODV are routing protocols for ad-hoc networks, with AODV for unicast traffic and MAODV for multicast traffic. MAODV allows each node in the network to send out multicast data packets, and the multicast data packets are broadcast when propagating along the multicast group tree.

- Message Formats of MAODV

Each multicast group has a unique multicast group address. According to the MAODV specification, each multicast group is organized using tree structure, composed of the group members and non group members. The nodes which are non group members that help in routing the data must exist in the tree to connect the group members. Associated with each multicast tree, the group member that first constructs the tree is the group leader for that tree, responsible for maintaining the group tree by periodically broadcasting Group-Hello (GRPH) messages in the whole network. The group leader also maintains the group sequence number, which is propagated in the network through the GRPH.

Each node in the network may maintain three tables.

- *Unicast Route Table* recording the next hop for routes to other destinations for unicast traffic.
- *Multicast Route Table*, listing the next hops for the tree structure of each multicast group. Each entry represents one group tree structure. Every node that belongs to that group tree should maintain such entries, with its own identity as group leader, group member, or router (non-multicast member that is in the tree to provide connectivity). Every next hop is associated with direction either downstream or upstream. If the next hop is one-hop nearer to the group leader, the direction is upstream; otherwise, the direction is downstream. The group leader has no upstream, while other nodes in the tree should have one and only one upstream.
- *Group Leader Table*. It records the currently-known multicast group address with its group leader address and the next hop towards that group leader when a node receives a periodic GRPH message. It includes the function of the *Request Table.*

Route Request (RREQ) Message Format is given in figure 1.

0 1 2 3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
Type J|R|G| Reserved | Hop Count |
| Other fields as specified for AODV.......

Figure 1. Route Request (RREQ) Message Format

Type → 1

J → Join flag; set when source node wants to join a multicast group.

R →Repair flag; set when a node wants to initiate a repair to connect two previously disconnected portions of the multicast tree.

Route Reply (RREP) Message Format is given in figure 2.

0 1 2 3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
| Type |R| Reserved |Prefix Sz| Hop Count |
| Other fields as specified for AODV.......

Figure 2. Route Reply (RREP) Message Format

Type → 2

R→Repair flag; set when a node is responding to a repair request to connect two previously disconnected portions of the multicast tree.

When the RREP is sent for a multicast destination, the Multicast Group Information extension is appended. Multicast Activation (MACT) Message Format is given in figure 3.

0 1 2 3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
| Type |J|P|G|U|R| Reserved | Hop Count |
| Multicast Group IP address |
| Source IP address |
| Source Sequence Number |

Figure 3. Multicast Activation (MACT) Message Format

MACT message contains the following fields:

Type → 4

J →Join flag; set when a node is joining the multicast group, as opposed to finding a route to the group for the transmission of data messages.

P→Prune flag; set when a node wishes to prune itself from the tree, unset when the node is activating a tree link.

G→Group Leader flag; set by a multicast tree member that fails to repair a multicast tree link breakage, and

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

indicates to the group member receiving the message that it should become the new multicast group leader.

U→Update flag; set when a multicast tree member has repaired a broken tree link and is now a new distance from the group leader.

R→Reboot flag; set when a node has just rebooted

Reserved→ Sent as 0; ignored on reception.

Hop Count→ The distance of the sending node from the multicast group leader, which is used only when the 'U' flag is set; otherwise sent as 0.

Multicast Group IP Address→The IP address of the Multicast Group for which a route is supplied.

Source IP Address→The IP address of the sending node.

Source Sequence Number→The current sequence number for route information generated by the source of the route request.

To prune itself from the tree (i.e., inactivate its last link to the multicast tree), a multicast tree member sends a MACT with the 'P' flag = 1 to its next hop on the multicast tree. A multicast tree member that has more than one next hop to the multicast tree should not prune itself from the multicast tree. Group Hello (GRPH) Message Format is given in figure 4.

0 1 2 3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
| Type |U|O| Reserved | Hop Count |
| Group Leader IP address |
| Multicast Group IP address |
| Multicast Group Sequence Number |

Figure 4. Group Hello (GRPH) Message Format

The format of the Group Hello message is illustrated above, and contains the following fields:

Type → 5

U→Update flag; set when there has been a change in group leader information.

O→Off_Mtree flag; set by a node receiving the group hello that is not on the multicast tree.

Reserved→Sent as 0; ignored on reception.

Hop Count→The number of hops the packet has traveled. Used by multicast tree nodes to update their distance from the group leader when the M flag is not set.

Group Leader IP Address→The IP address of the group leader.

Multicast Group IP Address→ The IP address of the Multicast Group for which the sequence number supplied.

Multicast Group Sequence Number→The current sequence number of the multicast group.

Control Messages

There are four types of *Route Requests*: RREQ, RREQ-J, RREQ-R and RREQ-JR.

RREQ is used under the following two situations:

1. Unicast route discovery and maintenance for reaching a specific node;
2. Unicast route discovery and maintenance for reaching a multicast group, when a node is not a multicast tree member but has multicast data packet(s) to send to that multicast group without knowing how to reach that tree.

RREQ-J is used under the following two situations:

1) When a node is not a multicast tree member but wants to join the multicast group;
2) Link breakage in the tree.

RREQ-R and RREQ-JR are used for tree merge.

Corresponding to different *Route Requests*, there are four different *Route Replies*: RREP, RREP-J, RREP-R and RREP-JR. The MACT messages are of three types: MACT-J, MACT-P and MACT-GL. MACT-J are used for tree construction when a non-member node wants to join the multicast group or when a link breakage is repaired in the tree. MACT-P is used for pruning a node from the tree if received from downstream. If received from upstream, MACT-P indicates not only pruning but also selecting a new group leader. MACT-GL is used for new group leader selection. The GRPH messages are of two types: GRPH, GRPH-U. GRPH is periodically sent out from the group leader in the whole network. GRPH-U is sent out from an upstream node to downstream nodes in the tree to change the group information. The one hop Neighbor-Hello message is used for detecting link failures in the proposed work.

B.SECURITY

Currently, MAODV does not specify any special security measures [9]. Route protocols, however, are prime

targets for impersonation attacks, and must be protected by use of authentication techniques involving generation of unforgeable and cryptographically strong message digests or digital signatures. In this work MAODV is modified by adding security to it. Apart from the normal encryption mechanism, symmetric encryption mechanism is adopted. This will help in saving the battery power up to some extend due to less computational complexity. The node's movements in MANETs change the topology frequently.

Group creation and group maintenance are very important in multicasting. The range which one node broadcast hello message to adjacent node is 2-hop. The hello message is to collect all information of nodes in the range of 2-hop. According to the information, the path is designed and groups will be constructed. The security is achieved by authenticating the groups. Due to the limited battery power of the nodes it is desirable to opt less computational methods for providing security.

Symmetric encryption method depicted in figure 5 is used for secure group communication [10]. One secret key is shared between the groups and using the secret key secure communication is achieved. For each group, one member, which handles the secret keys is called the key manager. A period called epoch by which the keys will be refreshed by the key manger for providing additional security. The change of key will be informed to all the group members by the key manager. In the case of the failure of the group leader another member will be the leader and handle the key mechanism. The encryption and decryption are done as follows.

$$Y = \mathrm{E}K(X) \quad (1)$$

$$X = \mathrm{D}K(Y) \quad (2)$$

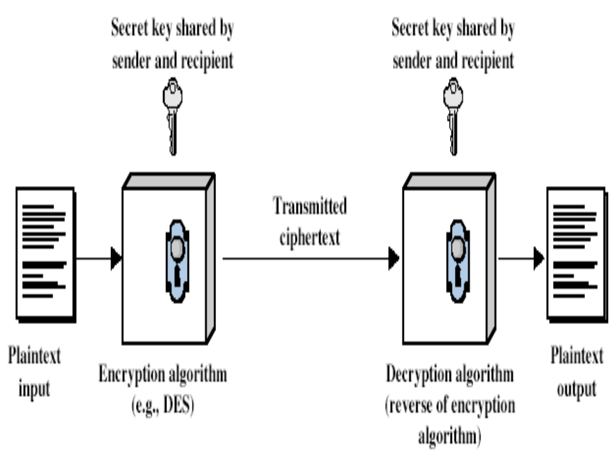


Figure 5. Symmetric Encryption

Equation (1) (on the sending side) represents the encrypted data using the secret key and the encryption algorithm. The encryption algorithm used is stream cipher. In (2), (on the receiving side) the encrypted data Y will be decrypted using the same key and the algorithm.

- Key Distribution

After the creation of the group, each group in the MANET shares a common key assigned by the key manager. The common key is assigned for the group is refreshed in each epoch by the key manager and that will be indicated to all group members. The new key will be issued by encrypting using the old key. The secret key will be ex-ored with the data and send by the multicast source. Using the same secret key the data will be decrypted by the destinations. The secure transmission is shown in figure 6.

The data need to encrypt by the multicast source will be split into block size of 'k' which is equivalent to the size

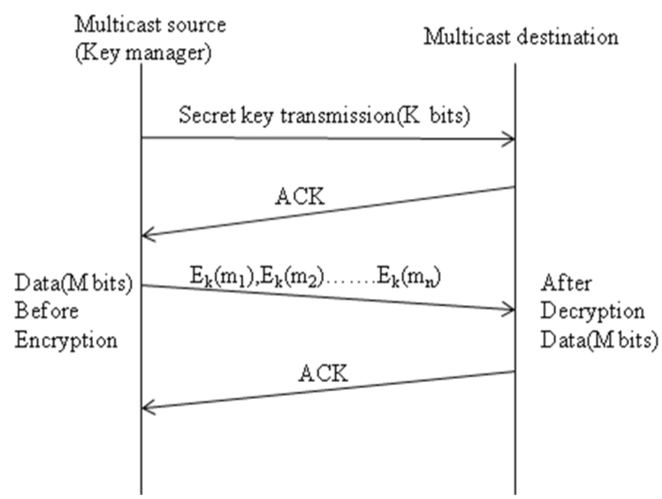


Figure 6. Secure transmission

of the key used for encryption. Each k bits are ex-ored with the selected secret key. The key set used for encryption is of equal size. When the last block of the data is not equal to the size of the key, then parity bits are added .In the multicast destinations reverse process will be performed. Confidentiality and integrity can be achieved by doing above steps. All the nodes should have enough memory space for storing the keys as and when required to act as the key manager.

Figure 7 contains a group of four members and one node is the key manager providing the secret key to the group members and after getting the key secure data transfer take places between the group members.. The network may have non group members also (shown in the figure).The non member cannot read the data even if it receives due to the security.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

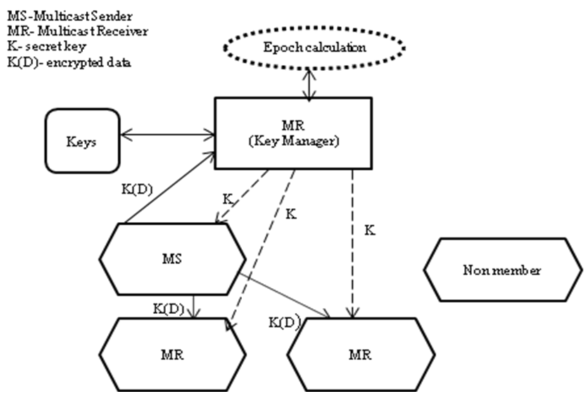


Figure 7. Secure Communication

C. QUALITY OF SERVICE

QoS mainly dea**ls** with bandwidth, delay, fault tolerance etc. The major intension of providing QoS is to efficiently utilize the available bandwidth by controlling the overhead. The protocol used here is MAODV which is an on demand routing protocol helps in reducing the control messages used for frequent routing updates. The symmetric key mechanism adopted in the above section will helps in reducing the average end to end delay due to less control messages. Fault tolerance is the main constraint considered in this work regarding QoS. In fault tolerance the node failures and link failures are included. Multipath routing is the solution used here to over the failures.

Fault Tolerance

Multipath routing protocols allow the establishment of multiple routes [12] between a single source and single destination node. This approach was initially developed to alleviate performance issues, as low throughput, low packet delivery ratio and high end-to-end delay, through redundant paths .Multipath routing protocols are attractive for improving reliability, load balancing, energy-conservation, and Quality-of-Service (QoS) [11].

Multipath routing consists of four main components: *route discovery*, *route maintenance*, *path selection* and *traffic allocation*. The route discovery and route maintenance find the multiple routes. A subset of these routes is chosen by the path selection component based on different criteria as path characteristics and interactions with the link layer. The traffic allocation strategy deals with how the data is distributed and sent through the selected paths. Each path is monitored and whenever it fails alternate path will be selected. The security of routing discovery is provided by the security mechanism integrated in the routing protocol. Two node-disjoint paths with the minimum sum of hops are selected by the source for reliability. Periodic updates of the routes are maintained by each node in the network. If no updates for a period of time the node is considered as failed or out of route. Alternate path will be selected for further data transmission. An alternate path for data transfer from source to destination is shown in figure 8.

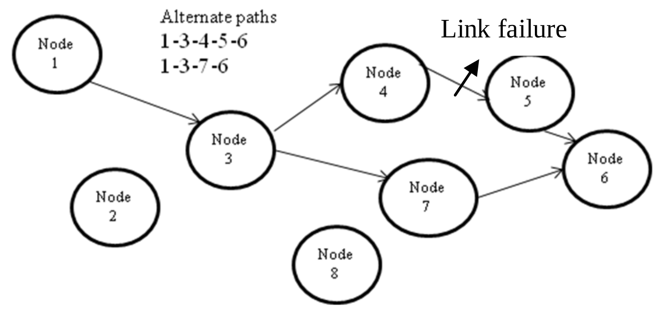


Figure 8. QoS by alternate paths

The sender of a multicast may move while transmitting or receiver may move while receiving the multicast message. Intermediate nodes will store the data while mobility happens and after finding the new path to the receiver the data will be forwarded. Alternate paths are chosen in order to provide better quality of service.

- Link Failure

In MANETs, the reliability of a path depends on the stability or availability of each link of this path because of the dynamic topology changes frequently. It supposes a free space propagation model [12], where the received signal strength solely depends on its distance to the transmitter. Therefore, using the motion parameters (such as speed, direction, and the communication distance) of two neighbors, the duration of time can be determined in order to estimate that two nodes remain connected or not. Suppose two nodes i and j are within the transmission distance r_a between them. Let (xi, yi) and (xj, yj) be the coordinate of mobile host i and mobile host j. Also let (vi,) be the speed and the moving direction of node i, let (vj, θj) be the speed and the moving direction of node j. The LET (Link Expiration Time) is predicted by [12] is calculated using (3)

$$LET= - (ab+cd) + \sqrt{(a^2+c^2)r_a^2-(ad-bc)^2} \;/ a^2+c^2 \qquad (3)$$

where $a = v_i\cos\theta i - v_j \cos\theta j$

$b = x_i - x_j$

$c = v_i\sin\theta i - v_j \sin\theta j$

$d = y_i - y_j$

Therefore, when vi = vj and θi =θj, LET tends to ∞. In other words, if LET is ∞, the link will remain connected at all times. On the other hand, if LET is negative, the link is disconnection In this way, the link existence can be calculated.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

- Node Failure

In MANET, mobile devices generally are dependent on finite battery sources. Once the battery power is completely consumed, then the mobile device will go down, that is the device is considered as under-failure. If the radio interface of the mobile device is not functioning, then all the communications from this device will be stopped. A prediction on node failure helps us in providing better QoS routing for ad hoc or sensor networks. One hop neighbor hello messages are used in order to detect the node failures. It will be send within the time interval. If there is no response for a period, the message will be send again. This will proceed until the number of attempts specified by the protocol is reached. If there is no response after the limit, alternate path will be selected and try for the same. Once again no response from the node, the node is assumed to be failed. The node which selected the failed node on its transmission path will go for alternate path .The protocol is on demand once the route is failed, then only will be searching for the new path. On a later stage the node is up can be detected using the one hop neighbor hello messages.

End to End Delay

For the path construction any metric can be chosen like path cost, path delay, path life time etc. In this work data path is evolved by considering the path delay, the time taken by the packet to flow from one node to another as well as the link expiration time.

The node number metric represents the path node number from the source to destination. For a path $P= (v_1, \ldots, v_n)$, the number of nodes is given in (4)

$$number\_node = |P| - 1 = n-1 \qquad (4)$$

Average end-to-end delay indicates the end-to-end delay experienced by packets from source to destination. The average end-to-end packet delay is computed as the ratio of total end-to-end delays to the aggregate number of packets successfully delivered to the destination nodes during a simulation run. The end to end delay is calculated using (5). Assume $p(s, d)$ denotes a path from the source to the destination, where $s \in N$ and $d \in (N - \{s\})$.Then the end to end delay of the whole path is defined as:

$$\text{Delay}(p(s, d)) = \sum_{e \in p(s, d)} \text{delay}(e) \qquad (5)$$

Throughput

Throughput can be expressed as the amount of data communicated from source node to destination node during a specified amount of time. Throughput calculation shown in (6)
.

$$\text{Throughput} = n\lambda \qquad (6)$$

where n is randomly selected source-destination pairs exchange traffic at rate λ .

IV. PERFORMANCE ANALYSIS

The performance evaluation of the protocol is carried out by using ns-2 simulator [13] [14][15]. The MAODV protocol is implemented in ns-2.The security and QoS are added to MAODV. The existing AODV protocol in ns-2 is upgraded for multicasting. The routines such as group creation, group deletion , group maintenance, multicast routing table, multicast node structure are included. Various timers are used for the group management. The group hello messages are used for QoS.

Nodes are deployed in an 1800 m X1800 m square area. The transmission range is 50 m. Simulation time is set to 52s. Bandwidth of the channel is set to $2X10^6$ Hz. The frequency assumed is 914MHz and the data rate is 2Mbps. Interface queue type used is CMUPriQueue.The performance evaluation done based on throughput and end to end delay. The transmission protocol used is TCP. Initial energy of the nodes are set to 3000J.Radio propagation is achieved using TwoRayGround model.FTP is build on top of TCP. Table 1 indicates the average end to end delay of randomly selected nodes from the multicast group of 5 members, table 2 indicates the average end to end delay of randomly selected nodes from the multicast group of 4 members and table 3 dictates the throughput comparison of the protocols. The graphical representations of the results are given in figures 9, 10 and 11 respectively.

Table 1. Average End to end delay of a group of 5 members

Members	MAODV	Modified MAODV
9	0.34	0.33
10	0.28	0.46
13	0.27	0.33
19	0.27	0.32

Table 2. Average End to end delay of a group of 4 members

Members	MAODV	Modified MAODV
5	0.28	0.33
7	0.26	0.36
8	0.30	0.32

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Table 3.Throughput

Granularity	MAODV	Modified MAODV
5	5.000175000	5.003021193
10	10.000255000	10.002580392
15	15.000155000	15.003834657
20	20.000355000	20.015930539
25	25.014710157	25.000334492
30	30.006501020	30.003123123
35	35.000269411	35.934717869

The result shows that the modified protocol having an improved throughput than MAODV. It also gives an almost consistent less delay while sending the packets.

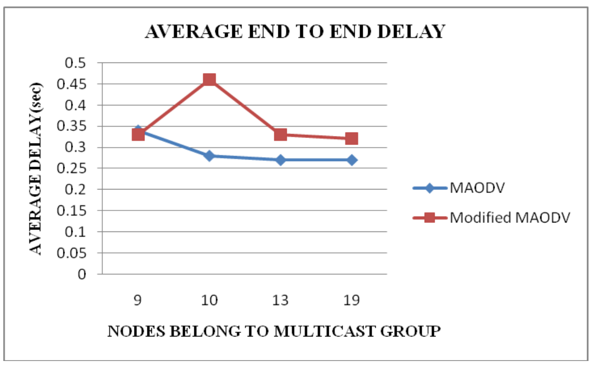


Figure 9. Average end to end delay of a group of 5

Figure 12 shows the average end to end delay when the node mobility happens. Throughput evaluation has been done by setting the granularity as five. The end to end delay has been calculated by randomly selecting the nodes from the same group. Two groups are considered for the average end to end delay calculation.

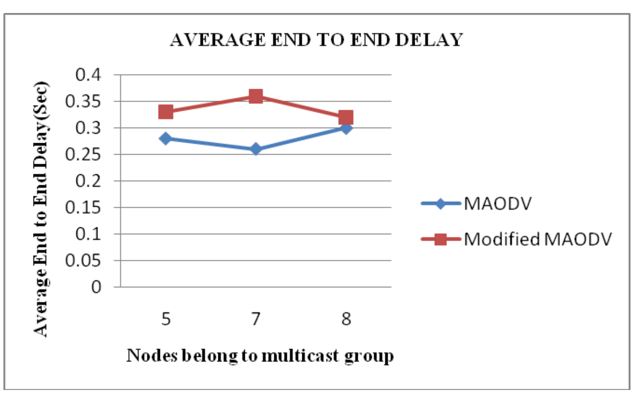


Figure 10. Average end to end delay of a group of 4

The groups are of 4 and 5 members each. Networks of 25 mobile nodes are considered for sample evaluation. The node numbers are ranging from 0...24. The performance can be even better by including proper quality of services.

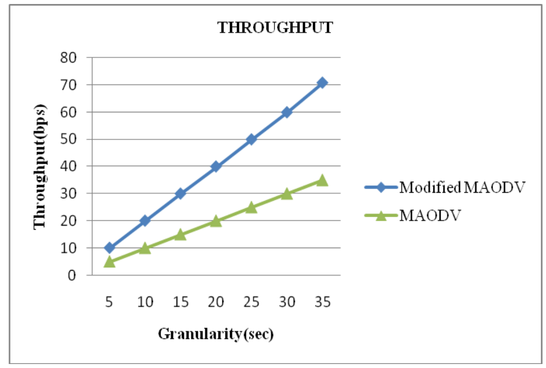


Figure 11. Throughput

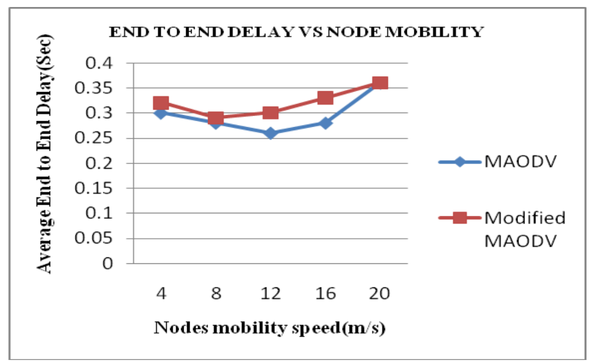


Figure 12. Average end to end delay vs. node's mobility speed

V. CONCLUSION AND FUTURE WORK

Application areas such as rescue sites, group conferences need the usage of multicasting. The major challenge facing in this area is the security. By incorporating the secret key mechanism the group communication is made secure. By adding the epoch concept the validity of the key is made even more secure. By including quality of service, the group communication is more efficient.

The results obtained shows that even though the overhead due to security is increased, not much affected the end to end delay and the throughput. By using the symmetric key mechanism which is less complex, the computational power needed is very minimum. The battery power of the node can be saved by this mechanism. The QoS includes the bandwidth, end to end delay and the fault tolerance such as node failures and link failures. The protocol can be modified efficiently in order to handle misbehaving nodes and selfish nodes.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

REFERENCES

[1] Sun B L.Long-life multicast routing protocol in MAODV based on entropy.Journal of computational information systems.2005,1(2):263-268.

[2].Royer, E. M. and Perkins, C. E.; "Multicast Ad hoc On-Demand Distance Vector (MAODV) Routing", IETF, Intemet Draft: draft- ietf-manet-maodv-00.txt, 2000.

[3] C.E. Perkins, E.M. Belding-Royer, and S.R. Das, “Ad Hoc On-Demand Distance Vector (AODV) Routing,” Internet Draft, draftietf-manet-aodv- 13.txt, Feb. 2003.

[4] ZhengMing Shen and Johnson P. Thomas , “Security and QoS Self-Optimization in Mobile Ad Hoc Networks”, IEEE transactions on mobile computing, vol. 7, no. 9, September 2008.

[5] S. Yi, P. Naldurg, and R. Kravets, “Security-Aware Ad Hoc Routing for Wireless Networks,” in Proc. International Conf. on Mobile Computing and Networking (MobiHoc 2001), Long Beach, CA, October 2001, pp. 299-302.

[6] Jiejun Kong, Yeng-zhong Lee, Mario Gerla, “Distributed Multicast Group Security Architecture for Mobile Ad Hoc Networks”.

[7] E. Crawley, R. Nair, B. Rajagopalan, and H. Sandick, "A Framework for QoS Based Routing in the Internet", August 1998, RFC 2386.

[8] R. Braden, D. Clark, and S. Shenker, "Integrated Services in the Internet Architecture: an overview", 1994, IETF RFC 1633.

[9] William Stallings, “Cryptography and Network Security: Principles and Practice”, 3-rd edition, Prentice Hall, 2003.

[10] H. Deng, et al. “Routing Security in Wireless Ad Hoc Networks,” *IEEE Communications Magazines*, vol. 40, no. 10, pp. 70 – 75,Oct. 2002.

[11] Shah H, Nahrstedt K. Predictive Location-Based QoS Routing in MANETs. In : Proceedings of IEEE International Conference on Communications (ICC2002),New York,April2002.

[12] William Su,Sung-Ju Lee, and Mario Gerla: Mobility Prediction in Wireless Networks.21st Century Military Communications Conference Proceedings. (MILCOM 2000). Los Angeles, CA, USA, Vol1, Oct (2000), pp.491-495.

[13] K.Fall and K.Varadhan, The ns-manual available at http: //www.isi.edu/nsnam/ns/.

[14] http://www.isi.edu/nsnam/ns/ns-documentation.html - ns Manual

[15] http://jan.netcomp.monash.edu.au/ProgrammingUnix /tcl/tcl_tut.html - Tcl tutorial

[16] Luo Junhai, Ye Danxia, Xue Liu, and Fan Mingyu, “A Survey of Multicast Routing Protocols for Mobile Ad-Hoc Networks”, IEEE communications surveys & tutorials, vol. 11, no. 1, first quarter 2009.

[17] Qifeng Lu, "A Survey on Vulnerability of Wireless Routing Protocols",Presentation, Virginia Polytechnic Institute and State University, July 2005.

[18] Thomas Kunz and Ed Cheng, "Multicasting in Ad-Hoc Networks: Comparing MAODV and ODMRP", Proceedings of the Workshop on Ad hoc Communications, Bonn, Germany, September 2001, pp. 16-21.

[19] Sencun Zhu, Sanjeev Setia, Shouhuai Xu, Sushil Jajodia, "GKMPAN:An Efficient Group Re-keying Scheme for Secure Multicast in Ad-Hoc Networks", First Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services (MobiQuitous'04), 2004

[20] Bing Wu, Jie Wu, Eduardo B. Fernandez Spyros Magliveras, "Secure and Efficient Key Management in Mobile Ad Hoc Networks", Proc. of the 1st Int'l Workshop on Systems and Network Security (SNS2005) (in conjunction with IPDPS), April 2005.

[21] R. Guerin and A. Orda, "QoS-based Routing in Networks with Inaccurate Information: Theory and Algorithms," Infocom'97, Japan,April 1997.

[22] C. E. Perkins, E. M. Royer, S. R. Das, "Quality of Service for Ad hoc On-Demand Distance Vector Routing," draft-Perkins-manet-aodvqos-02.txt, IETF Internet Draft, work in progress, October 2003.

[23] Yih-Chun Hu, David B. Johnson, "Securing Quality-of-Service Route Discovery in On-Demand Routing for Ad Hoc Networks," Conference on Computer and Communications Security Proceeding of the 2nd ACM workshop on Security of ad hoc and sensor networks. Washington DC, USA, SESSION: Secure routing in ad hoc networks, Pages: 106- 117,2004, ISBN:1 -58113 – 972- 1.

[24] H. Xiao, W.G. Seah, A. Lo, K.C. Chua, "A Flexible Quality of Service Model for Mobile Ad-hoc Networks (FQMM)", in Proceedings of IEEE Vehicular Technology Conference (VTC 2000-Fall), Vol. 1, No.4, May 2000, pp.397-413.

[25] William Stallings, “Cryptography and Network Security: Principles and Practice”, 3-rd edition, Prentice Hall, 2003.

(IJCSIS) International Journal of Computer Science and Information Security,
Vol. 8, No. 4, July 2010

Analytical Comparison of Fairness Principles for Resource Sharing in Packet-Based Communication Networks

Yaser Miaji and Suhaidi Hassan

InterNetWorks Research Group, UUM College of Arts and Sciences
Universiti Utara Malaysia, 06010 UUM Sintok, MALAYSIA
ymiaji@internetworks.my suhaidi@ieee.org

ABSTRACT

Current Internet users are enormously increased and application that they are using is magnificently bandwidth devoured. With this manner, Internet is no longer a fair and protective environment. The diversity in the Internet applications required a reconsideration of the mechanisms used to deliver each packet pass through a router in order to provide better fairness and more protective place. Furthermore, the observer of the Internet packet could easily identify the purpose of the delay which is indeed caused by the queuing in the output buffer of the router.

Therefore, to reduce such delay for those sensitive applications such as real-time applications, scholars develop many fairness principle which by turn could improve the QoS and hence the fairness and the protection aspect. This study highlight most famous fairness principles used in the literature and some other novel ideas in the concept of fairness. The analytical comparison of these principles shows the weakness and the strength of each principle. Furthermore, it illuminates which fairness principle is more appropriate in which environment.

Keywords-components; Fairness, max-min, proportional fairness, balanced, max-min charge

1. INTRODUCTION

Internet utilization in public and private sector is magnificently growing with extraordinary manner. The occupation of the World Wide Web is unpredictable over time frame. Daily usage of the Internet resources with current scrambles in network access is hard to be estimated and hence the distribution of these resources is dynamic. This dynamic behavior leads to vagueness in constructing the essential principle of fairness for resource utilization.

Furthermore, not only the dynamic attitude of the resource utilization is an issue, the behavior and the characteristics of the application itself also, play a potential responsibility in structuring the fairness principle. Some applications require more sensitive pamper and care such as voice and interactive application such as video conversation and so forth. The sensitivity of these applications significantly involved in fairness principle.

Moreover, providing Quality of Service (QoS) is one big dimension which should be achieved if not fully at least to the large extent. QoS requirements rhyme heavily with user and application requirements. Even though, Service Providers (SP) is one potential dimension which tighten fairness principle, their requirements is highly depend on financial matters.

Fairness principle is indeed, applied in routers or to be more specific in the process of scheduling the transmission of the packets over a shared link. Fairness principle should provide three primary function selection, promptness, and QoS consideration. Selection is the basically which packet deserves to be transmitted. Promptness means when the selected packet will be transmitted. QoS requires considering the delay, loss and error of overall network performance.

Scholars, since the discovery of the sensitive and bandwidth hanger applications, dedicate their research in providing superior fairness and larger protection for these applications over others less sensitive. This paper demonstrates most available and used fairness principles in scheduling packets depending on application sensitivity and user usage. The rest of the paper is organized as following. Next section gives the state diagram of the literature and brief information about the evolution of the fairness principle. This is followed by thorough conceptual and analytical illustration of five fairness definition namely; max-min fairness, proportional fairness, utility fairness, balanced fairness, and max-min charge fairness. Section four compares and contrasts all six principles and finally the conclusion and future works are drawn.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

2. Mind-Map of Fairness Literature

In this section, related works to the fairness is presented in state diagram or min-map diagram to correlate and track the evolution of fairness principle. Exhibit 2.1 shows the mind-map diagram which explains the evolution of fairness principle. In 1967, Kleinrock [1] published his article in sharing one common resource. Although the article is primarily designed for addressing this specific issue from processor sharing prospective, it opened sites in discussing fairness in networks since process sharing environment shares some similarity with resource sharing in the Internet or networks. Kleinrock then wrote his book which consists of two volume in queuing systems [2, 3]. In this book the essential ideas and explanation of max-min fairness principle is been demonstrated with the aid of mathematic. Jaffe [4] incorporates the max-min fairness principle explicitly in network resource sharing. This concept is been presented in data networks book written by Bertsekas et al. [5].

Nevertheless, the concept and regulations which rule max-min fairness and lead to its result are not convenience and does not provide the efficient fairness from Kelly point of view [6, 7]. Consequently, he proposed an alternative fairness principle named as proportional fairness. This concepts is further developed by Massoulie and Roberts [8]. Bothe principles; max-min and proportional are further compared and thoroughly analyzed by Denda et al. [9]. However, the advocates of proportional fairness has comprehensively illustrate the principle in [10].

Despite the success of the most famous principles; max-min and proportional, they have some weaknesses which are discovered by Bonald and Proutiere [11]. Balanced fairness is their proposal which is inspired by Erlang [12] ideas, has different approaches. All three principles; max-min, proportional and balanced fairness are presented in Bonland et al. paper [13]. Bonland has provided some comparison using analytical demonstration. Another fairness view is called utility fairness introduced by Cao and Zegura [14]. Utility fairness has adopted the concept of utility proposed in [15]. All the above mentioned fairness definition have been presented in [16] by Hosaagrahara.

However, these four principles; max-min, proportional, balanced and utility fairness are in principle correlated and based on bandwidth allocation with different approaches in determining the proper algorithm to chose the next packet in line. The entire principle of bandwidth allocation has been criticized in Briscoe article [17]. Therefore, Miaji and Hassan in [18] proposed a new vision of fairness by providing the principle of charge allocation rather than bandwidth allocation and it named as max-min charge. Max-min charge is a new fairness principle based on charge allocation instead of conventional bandwidth allocation. Next section presents all the above mentioned five fairness principles conceptually and analytically.

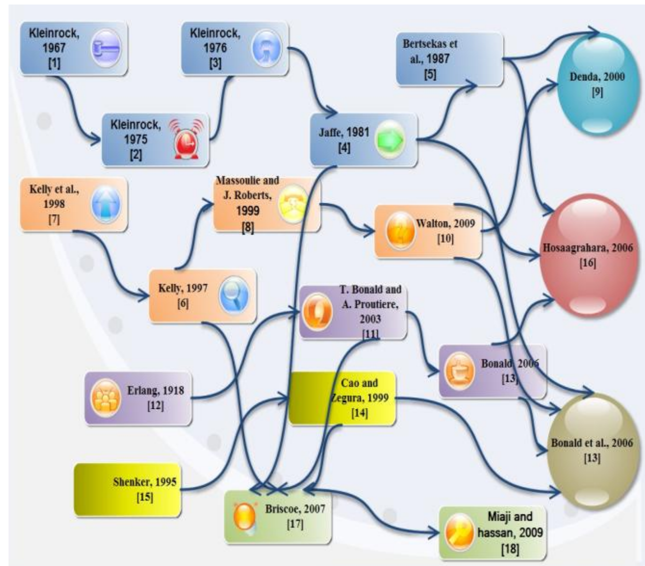


Exhibit 2.1: Mind-map literature of fairness principles

3. Principles of Fairness

Approaching an optimum fairness in shared elastic environment such as the Internet is complicated and frustrated. As a consequence, different proposals have been drawn to accomplish the mission in several prospective. This section provides rigorous knowledge in the most five adopted fairness notions. This comprehensive illustration will reach the conceptual and analytical approach of each o these five notions. Next section compares and contrasts these five principles.

Before the explanation of the five notions mentioned earlier, a scenario of shared resource is been assumed. So, let consider the following scenario. Consider a contended user n with demands $D = \{ d_1, d_2, \dots \dots \dots .. d_n\}$ varies from one user to another. Those users are sharing the one resource R. Additionally, each user is allocated a specific portion $A =$ Error! Bookmark not defined. of the resource R according to a policy P. There are two main stipulations for such allocation;

a. The resource which is allocated is finite and limited.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

b.There is no resource feedback from users' side.

Consequently, any policy abides by these two conditions is said to be active and defined as follows [16]:

Definition 3.1: The policy P is said to be active if, for all possible demands D, it results in an allocation A such that:

1. $a_i = 0, for\ all\ i\ \in \{1,2, \ldots .n\}$
2. $\sum_{i=1}^{n} a_i\ \le R$

Now, let establish the investigation in the five fairness principles.

3.1 Max-min Fairness

Let first simplify the principle of ma-min fairness be the following example. Let assume that there are buckets which are corresponding to the demand d_i of the users. Moreover, let assume that all buckets share the same tab which corresponds to the resource R. Therefore, since the resource is limited and the buckets cannot, indeed, provide any resource enhancement which there is no other resource except the one which is shared as seen in exhibit 3.1.

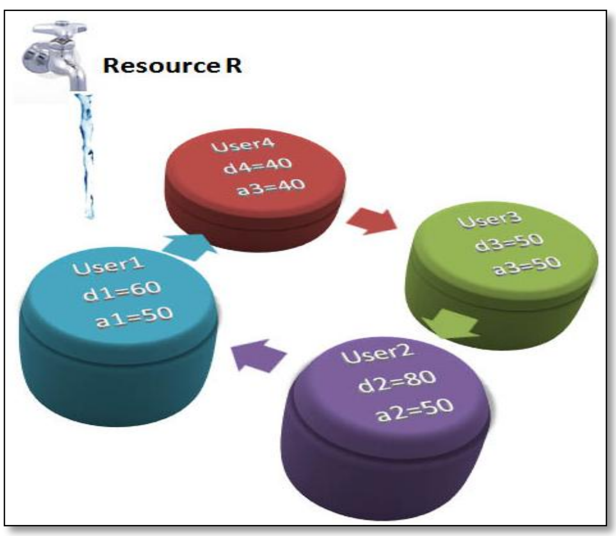


Exhibit 3.1: Users Share the same resource

According to max-min principle no user will obtain more than its demand and also, all not fully served users will be equally allocated in term of the resource.

Therefore, user 1, 2, and 3 will take exactly what they demand since their demands is the lowest. In comparison, user 4 and 5 will take equal resource allocation no matter what they demand for.

Additionally, any user attempts to increase its allocation will result in decrease in the resource allocated to another. Furthermore, it could be obviously seen that the attempts to increase the demand will not influence the decision of allocation [16].

Exhibit 3.1 provides us with much information which has not been illustrated yet. The essential inspiration of max-min fairness is the Pareto superiority as well as Pareto efficiency which were suggested by Pareto [19, 20]. In fact, Pareto proposed his notion in political economic and it has two main concept; superiority and efficiency for two active allocation. Firstly, if we have to allocate $a_1 and\ a_2$ to two different resources $u_1 and\ u_2$, a_2 is considered as Pareto superior with respect to a_1 if a_2 expands the allocation of at least one entity while not reducing the allocation of any other entity; for instance, at least one user prefers a_1 over a_2. In the case of exhibit 3.1, user 4 prefers to obtain 40 units over 50 units and no other user request it. This preference will affect other users [21].

Secondly, an allocation is considered as Pareto optimal if it is active and Pareto superior to all other active allocations. Indeed, Max-Min fairness shows its Pareto optimality and hence it is unique since it is the only notion which meets the conditions of the Pareto optimality [22].

Now, let take the analytical vision of the notion of Max-Min fairness. So, let presume that a_1 is the allocation dedicated for u_1 with demand d_1 in flow i and a_2 is the allocation specified for u_2 with demand d_2. If we assume the $d_1 = d_2$ then the following theorem could be deduced;

Theorem 3.1:

The Max-Min fairness is unique.

Proof:

Let $u_1$and u_2 two users with demands $d_1$and d_2 respectively and the resource allocated for them is $a_1\ and\ a_2$ respectively as well. So, if $d_1 = d_2$ then the allocation results could be;

$$a_1 = a_2, a_1 > d_2\ or\ a_1 < a_2$$

Only first one is possible since the remaining two are not Max-Min fair.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Moreover, consider $r_1^{'}$ is the service received by u_1, then if $r_1^{'} > d_1$ that means this allocation is not Max-Min fair because in Max-Min fair the following should be true: $r_1^{'} \leq d_1$. Hence the following definition is true for Max-Min fair:

Definition 3.1:

A policy P is considered Max-Min fair if and only if satisfies the following conditions [22, 23]:

1- A is active;
2- Any attempts to increase and allocation for specific user result in a decrease in another user with equal or less value.

Therefore, Max-Min policy should have the following properties;

Property 3.1: No user gets resource allocation than what it have been requested.

Property 3.2: users with same demands will be allocated similar resource.

Property 3.3: Any increase in the demand will not affect the allocation procedure.

3.2 PROPORTIONAL FAIRNESS

The idea of the proportional fairness is, indeed, proposed after the discovery of some gap in the fairness of Max-Min. We will simplify this concept by illustrating a wireless node example [10].

It well known that the fairness goal is not to maximize the overall throughout or the bit rate or increase the efficiency, it rather to be fair in allocating the bandwidth in accordance to the current network status. From this sense, consider a constant[1] wireless network where there are two status of a node either good or bad. Therefore, in order to achieve high throughput and hence to maximize the bit rate or increase the efficiency, it is better to allocate more bandwidth, transmission power and so on to those good nodes since the bad one will experience more loss and required more bandwidth with less throughput [13].

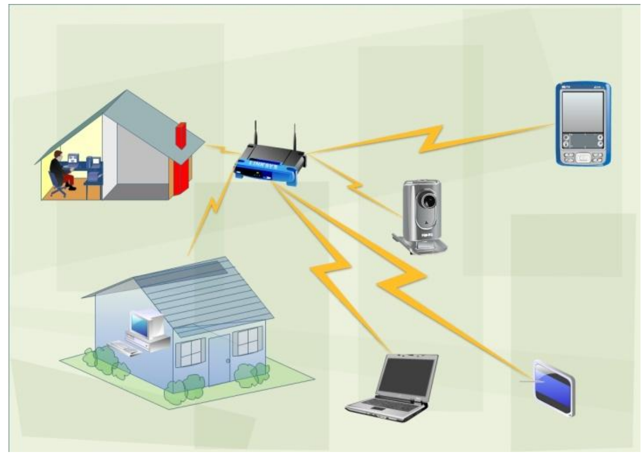


Exhibit 3.2 a: Max-Min fairness

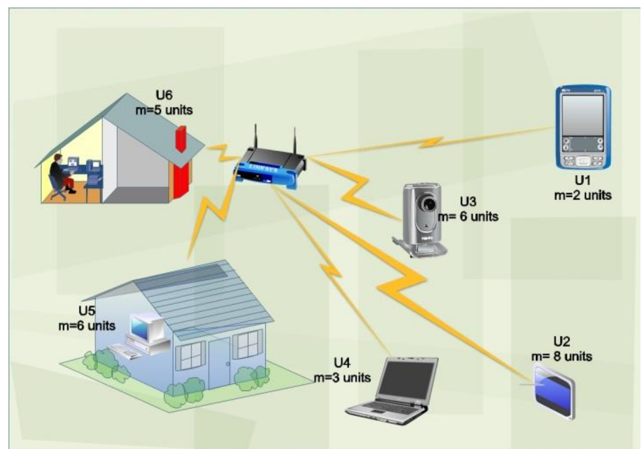


Exhibit 3.2 b: Proportional fairness[2]

Nevertheless, this maximization will not be fair since those nodes with bad radio channel will suffer from starvation. In Max-Min concept as shown in exhibit 3.2a, those nodes with bad radio channel will be allocated more bandwidth since the main aim of such principle is to maximize the minimum. However, from proportional fairness point of view, this solution is not the optimum.

Therefore, there is a trade of between efficiency and fairness. Proportional fairness is trying to solve and hence minimize this trade of by proposing the concept of allocating bandwidth in proportion to charge [6, 7].

[1] This situation is likely to be impossible especially in the case of mobile wireless environment.

[2] The width of the wireless communication link corresponds to the bandwidth allocated to this specific user.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Logarithmic approach has been dedicated to such approach. So, proportional fairness concept proposed the notion of price per unit used or shared (see exhibit 3.2b). If we assume that user u_1 is charge of an amount of m_{u1} for unit shared. Therefore, in proportion to m_{u1} this user will be allocated a_{u1}. As a consequence the problem of maximization could be formed as following;

Maximize $\sum_{u1} m_{u1} \, log a_{u1}$

So, the allocation for each user is depend on the amount it is charged. This gives some restriction in the utilization of such concept which will be discussed later in the analysis and comparison section.

3.3 Utility Fairness

The concept of utility fairness is easily to be inferred from its name. This notion is based on the utility or the application. It basically, derives the bandwidth allocation in accordance to the characteristic of the application to be transmitted through the link. Therefore, packets which has elastic or more tolerance in term of delay or loss or any other specified criterion, are allocated bandwidth depending on its specifications, behavior, and characteristics [14].

Therefore, in the case of the identical utility or packet specification or in other words applications, packets will be treated as in Max-Min fairness. On the other hand, as the application or packets diverse in its characteristics or manners, the allocation scheme will also, changed and is highly depends on the utility.

To simplify the idea of utility life example is been provided. Now, consider an apple which needs to be divided among three people fairly as in exhibit 3.3. The simple and basic way is to allocate one third of this apple to each person equally as shown in exhibit 3.3a. However, this sort of division is considered unfair if the circumstances of the people are not equal.

So, now consider the first person is a child how will any way, cannot eat more than a quarter of the apple. The second person is in diet and he also, cannot eat more than a quarter of the apple and the third is very hungry energetic youth. Consequently, according to the utility as one half is allocated to the youth, quarter for the child and the last quarter portion is allocated for the person in diet (see exhibit 3.3b).

(a) (b)

Exhibit 3.3: example of utility fairness

Cao [14] in his article proposed and proof the following theorem;

If a_1 is the allocated bandwidth for i session, C is the link capacity, the real utility function for session i is g_i , e_i is the error in the advertised bandwidth and k_i is the difference between the utility achieved by session i and the allocation deserved by the same session then;

$$|\Delta k_i| \leq 2 \, max_i \, (|e_i|)$$

A quantitative measure of the error in utility allocation is given by such theorem which resulted from the inaccurate information. Moreover, it reveals that there is a strong relationship between the error of utility allocated to an individual source and the accuracy of advertised utility functions; nevertheless, it is not affected by the number of sources sharing the same bottleneck link and hence no harms from any exponential increase in the users side.

3.4 Balanced fairness

The proper definition of balanced fairness is the unique balanced allocation such that $\emptyset(x)$ belongs to the boundary f the capacity set in any state $x \neq 0$. If Φ corresponds to the balance function, the following equation is true in any state $x \neq 0$.

$$\emptyset_i(x) = \frac{\Phi\,(x-e_i)}{\Phi(x)}\ ,\ i = 1,2, \dots . N \qquad (3.1)$$

Therefore, $\Phi(x)$ is recursively defined as the minimum positive constant β such that the vector $(\Phi(x-e_1, \dots, \Phi(x-e_n))/\beta$ belongs to C .

Balanced fairness is a new notion of bandwidth allocation with the very gratifying property that flow level performance metrics are insensitive to detailed traffic

characteristics [24]. This is particularly important for data network engineering since performance can be predicted from an estimate of overall traffic volume alone and is independent of changes in the mix of user applications [13].

3.5 Max-Min charge

Max-Min charge has taken new different vision of fairness in packet switching networks. The authors claim that to provide better fairness and proper protection to any user in a common shared resource, some aggressive penalty should applied for those who are maliciously use the sharing procedure [18].

Let take the analogy of multiple buckets sharing one fountain or resource as in exhibit 3.4. So, let consider that u_1 greedily attempts to gain more bandwidth by initiating several session with multiple connation and hence reserves more bandwidth than the others. Such manner could breaches both the protection of other users who indeed fairly be using the resource and the fairness by making u_1 get double service than the others.

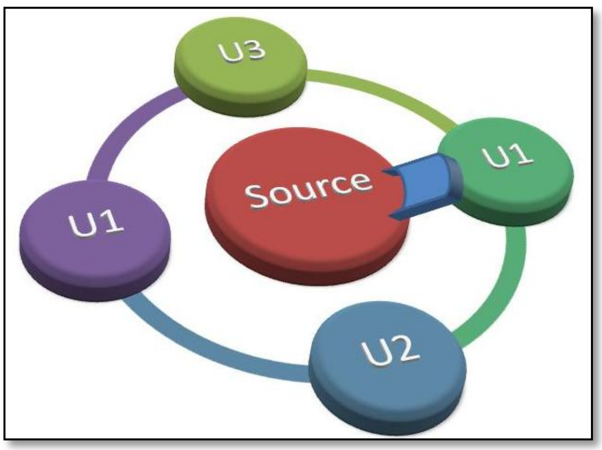


Exhibit 3.4: analogy of max-min charge

Nevertheless, Max-Min fairness has nothing to do regarding such issue since it concern about fairness among flows and not users. However, Max-Min charge assigns a specific values Usr_{le} and parameters to each user. The following equation has been deducted to improve the protection level:

$$F_i^n = S_i^n + \frac{L_i^n + Usr_{le}}{\emptyset_i} \qquad (3.2)$$

By this notion, the only user who will suffer from any increase in the demand or in queue number is the misbehaved user. Some charges will be applied to such users and hence minimize the allocation.

4. Comparison of Fairness Principles

The current status of the Internet provides only best-effort service. Consequently, providing enhancement for traffic flows for bandwidth reservation purposes is almost absent, or to be more precise bounding delay and jitter is not up to the expectation level or even not met. Moreover, any further modification in the protocols to be able to adopt the concept of reservation high efficiency of Quality of Service (QoS) required a crucial modification in the core of the Internet which is unachievable. These boundaries rigorously reduce the ability of flows to demand guarantees from the Internet, and the capability of the Internet to put forward and accomplish such guarantees.

If these constraints taken in account, the most appropriate notion to be considered is max-min fairness. The principle of proportional fairness necessitates flows to transmit information about their bandwidth requirements and reservations to each router along their rout. The principle of utility fairness is unclear in term of the specification of the utility function and it rather demands flows to convey their utility.

Nevertheless, minimum information about the flows among all notions is required by the principle of max-min fairness; a flow has a demand of unity if it has a packet enqueued and has a demand of zero otherwise. This information is, indeed, promptly available to each router and therefore the max-min principle of fairness is the most amenable to implementation. Likewise, max-min fairness is presently the most accepted principle of fairness in the research community.

5. Conclusion

The nature of the Internet traffics is random and dynamic; therefore, such behaviors should be taken in account once the issue of resource allocation is investigated. It has proven that max-min fairness, proportional fairness and balanced fairness provide stability to the network particularly when the vector of traffic intensities depends on the interior of the capacity set. It is also, proven that balance property have not been met by max-min fairness notion with the exception of the trivial case where the network condenses to a set of independent links. This justifies the limitation of the analytical results for this allocation and strengthens the assumption that such results should not be excluded. Proportional fairness is not balanced either except in some specific cases.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Balanced fairness, conversely, is balanced by construction and consequently leads to a well-mannered queueing network. Nevertheless, it is lack of practical implementation; nonetheless, balanced fairness could to a certain extent be contemplated as a mathematical tool practical for the performance evaluation of more practical allocations like max-min fairness and proportional fairness.

Max-min fairness, in contrast, could achieve much worse performance than balanced fairness and proportional fairness. This statement is been drown since max-min fairness is giving the absolute priority to flows with small bit rates. Therefore, in wireless communication, such notion results in an inefficient allocation where flows that experience bad radio conditions expend most radio resources. Proportional fairness and balanced fairness, alternatively, are homothetic as a result the allocated network resources will not rely on the radio conditions. Thus, in heterogeneous networks, such allocations are much more client and robust than max-min.

Max-Min charge is new definition which required more conceptual and analytical proof to take place in the competition. However, its primary ideas demonstrate its novelty. Finally, market action is required to improve these less-developed notions which could result in enormous improvement in the fairness and protection of the scheduling.

REFERENCES

[1] L. Kleinrock, "Time-shared Systems: a theoretical treatment," Journal of the ACM (JACM), vol. 14, pp. 242-261, 1967.

[2] L. Kleinrock, "Queueing systems, volume 1: theory," John Wiley & Sons, 1975.

[3] L. Kleinrock, Queueing Systems: Volume 2: Computer Applications: John Wiley & Sons New York, 1976.

[4] J. Jaffe, "Bottleneck flow control," Communications, IEEE Transactions on [legacy, pre-1988], vol. 29, pp. 954-962, 1981.

[5] D. P. Bertsekas, R. Gallager, and T. Nemetz, Data networks: Prentice-hall Englewood Cliffs, NJ, 1987.

[6] F. P. Kelly, "Charging and rate control for elastic traffic," European transactions on Telecommunications, vol. 8, pp. 33-38, 1997.

[7] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, "Rate control for communication networks: shadow prices, proportional fairness and stability," Journal of the Operational Research society, vol. 49, pp. 237-252, 1998.

[8] L. Massoulie and J. Roberts, "Bandwidth sharing: objectives and algorithms," 1999.

[9] R. Denda, A. Banchs, and W. Effelsberg, "The fairness challenge in computer networks," 2000, pp. 208-220.

[10] N. S. Walton, "Proportional fairness and its relationship with multi-class queueing networks," The Annals of Applied Probability, vol. 19, pp. 2301-2333, 2009.

[11] T. Bonald and A. Proutiere, "Insensitive bandwidth sharing in data networks," Queueing systems, vol. 44, pp. 69-100, 2003.

[12] A. K. Erlang, "Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges," The Post Office Electrical Engineers' Journal, vol. 10, pp. 189–197, 1918.

[13] T. Bonald, L. Massoulié, A. Proutiere, and J. Virtamo, "A queueing analysis of max-min fairness, proportional fairness and balanced fairness," Queueing systems, vol. 53, pp. 65-84, 2006.

[14] Z. Cao and E. W. Zegura, "Utility max-min: an application-oriented bandwidth allocationscheme," 1999.

[15] S. Shenker, "Fundamental design issues for the future Internet," Selected Areas in Communications, IEEE Journal on, vol. 13, pp. 1176-1188, 1995.

[16] M. Hosaagrahara, "A generalized framework for achieving max-min fairness: theory and applications." vol. PhD: Drexel University, 2006, p. 134.

[17] B. Briscoe, "Flow rate fairness: dismantling a religion," ACM SIGCOMM Computer Communication Review, vol. 37, p. 74, 2007.

[18] Y. Miaji and S. Hassan, "Just Queueing (JQ): Scheduling Algorithm for the Internet," in The First International Conference on Networks & Communications (NetCoM-2009), 2009, pp. 161-165.

[19] V. Pareto, "The new theories of economics," The Journal of Political Economy, pp. 485-502, 1897.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[20] V. Pareto, R. Marchionatti, and F. Mornati, Considerations on the fundamental principles of pure political economy: Routledge, 2007.

[21] E. Karipidis, N. D. Sidiropoulos, and Z. Q. Luo, "Quality of service and max-min fair transmit beamforming to multiple cochannel multicast groups," IEEE Transactions on Signal Processing, vol. 56, pp. 1268-1279, 2008.

[22] D. Chakrabarty, J. Chuzhoy, and S. Khanna, "On allocating goods to maximize fairness," 2009, pp. 107-116.

[23] A. Sridharan and B. Krishnamachari, "Maximizing network utilization with max–min fairness in wireless sensor networks," Wireless Networks, vol. 15, pp. 585-600, 2009.

[24] T. Bonald, A. Proutiere, J. Roberts, and J. Virtamo, "Computational aspects of balanced fairness," 2003, pp. 801–810.

AUTHORS PROFILE

Yaser Miaji received the B.E. form Riyadh College of Technology, Saudi Arabia and M.E. degrees, from University of New South Wales, Australia. in 1997 and 2007, respectively. He is [16]currently a doctoral researcher in Computer Science in the University Utara Malaysia. Previously, he works as a lecturer in the College of Telecommunication and Electronic in Jeddah from 1998-2206. His research interest includes digital electronics, computer network, distributed system and genetic algorithm. He is a member of InternetWorks research group, IEEE, ACM ISOC and STMPE.

Suhaidi Hassan PhD SMIEEE is an associate professor in computer systems and communication networks and the Assistant Vice Chancellor of the Universiti Utara Malaysia’s College of Arts and Sciences. He received his PhD in Computing from University of Leeds in United Kingdom, MS in Information Science from University of Pittsburgh, PA and BS in Computer Science from Binghamton University, NY. He currently heads the InterNetWorks Research Group at the Universiti Utara Malaysia and chairs SIG InterNetWorks of the Internet Society Malaysia Chapter.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Multiple Values Bidirectional Square Root Search

Syed Zaki Hassan Kazmi
Department of Computer Science
IQRA University Islamabad Campus, Pakistan
zaki.mzd@gmail.com

Syeda Shehla Kazmi
Department of Computing & Mathematics
Manchester Metropolitan University, United Kingdom
shehla_kazmi@hotmail.com

Jamil Ahmad
Department of Computer Science
IQRA University Islamabad Campus, Pakistan
jamil@iqraisb.edu.pk

Syeda Sobia Hassan Kazmi
Department of Computer Science
University Of Azad Jammu And Kashmir
Muzaffarabad A.K, Pakistan
fajar_zara@hotmail.com

***Abstract*—The research in hand is an effort to introduce a new efficient searching technique known as Multiple Values Bidirectional Square Root Search. In this technique, a sorted list of values can be searched from another sorted list very efficiently. The overall time for sorting the values to be searched and searching is less than the time taken by the Linear Search and Binary Search. In this technique, the size of targeting list is reduced for every next value to be searched. As a result the searching time of remaining values to be searched is reduced, while in linear search and binary search the size of target list remains same for every value to be searched. In other words, we can say that the targeting list is traversed completely each time for every value to be searched from beginning to end. Running Cost analysis and the result obtained after implementation are provided in the graphical form with an objective to compare the efficiency of the technique with linear search as well as binary search, as binary search is consider as an efficient and fast one.**

Keywords-Searching; Linear Search; Binary Search.

I. INTRODUCTION

Algorithms have a vital and key role in solving the computational problems, informally an algorithm is a well defined computational procedure that takes input and produces output. Algorithm is a tool or a sequence of steps to solve the computational problems [1].

In computer science, a searching algorithm, broadly speaking, is an algorithm that finds a particular data from a data set. Searching takes a problem as input and returns a solution to the problem, usually after evaluating a number of possible solutions. Efficient searching is important to optimizing the use of other algorithms (such as sorting algorithms) that require searching the location of new item in sorted list such as in insertion sort. There are a lot of searching techniques, currently used in industry and academia, to find the data of various forms and from different areas.

The study in hand proposes a new searching technique that is tested and analyzed against binary search to provide its efficiency. Searching is of considerable importance as the human is possessed in searching the information/knowledge. To search the information efficiently the efficient searching technique is very important. To facilitate the human, computers consume a substantial time in searching the data.

This paper is organized as follows; Section II presents a brief review of existing searching algorithms. Section III presents the description of proposed solution. Section IV presents the proposed algorithm. Section V presents Running Cost Analysis. Section VI present comparison between multiple values bidirectional square root search with linear search and binary search. Section VII ends with concluding remarks.

II. A BRIEF REVIEW OF EXISTING SEARCHING ALGORITHMS

A number of searching techniques are currently used in the field of computer science. This section will briefly discuss some of the trendy searching techniques among them. These are following:

A. Linear Search

It is the simplest method to find the particular element in the list. It checks all of the elements in the list and particularly checks in sequence and one at a time to reach the particular value. It is the special case of Brute Force Search. Its worst case cost is proportional to the number of elements in the list. It is O (n) [9].

B. Binary Search

"In computer science, a binary search is an algorithm for locating the position of an element in a sorted list. It inspects the middle element of the sorted list: if equal to the sought value, then the position has been found; otherwise, the upper half or lower half is chosen for further searching based on whether the sought value is greater than or less than the middle element. The method reduces the number of elements needed to be checked by a factor of two each time,

and finds the sought value if it exists in the list or if not determines "not present", in logarithmic time. A binary search is a dichotomic divide and conquer search algorithm. Its running time complexity is O(logn)"[5].

III. Multiple Values Bidirectional Square Root Search

In this technique, a sorted list of values can be searched from another sorted list very efficiently. The overall time for sorting the values to be searched and searching is less than the time taken by the Linear Search and Binary Search. In this technique, the size of targeting list is reduced for every next value to be searched. As a result the searching time of remaining values to be searched is reduced, while in linear search and binary search the size of target list remains same for every value to be searched. In other words, we can say that the targeting list is traversed completely each time for every value to be searched from beginning to end. The steps of the proposed Multiple Values Bidirectional Square Root Search algorithm are as follows:

1. Take 1st element from the item list and assign it to "item"
2. Assign 1st element of the Data list to left pointer and last element of the Data list to right pointer.
3. Take square root of list length and assign to "sqr".
4. If value on the left pointer is less then item ,increment left pointer .The new location of left pointer is getting by adding left pointer current location with "sqr".
5. If value on the left pointer is greater then item, change the location of right pointer by calculating new location as subtracting 1 from current location of left pointer. Also change the location of left pointer by subtracting left pointer current location with "sqr" and increment by 1. The new value of "sqr" is calculated by taking square root of "sqr".
6. Similarly If value on the right pointer is greater then item, decrement right pointer .the new location of right pointer is getting by subtracting right pointer current location with "sqr".
7. If value on the right pointer is less then item, change the location of left pointer by calculating new location as adding 1 with current location of right pointer. Also change the location of right pointer by adding right pointer current location with "sqr" and decrement by 1. The new value of "sqr" is calculated by taking square root of "sqr".
8. Repeat step 3 to 6 until value of left pointer or right pointer match with item, or the value of "sqr" become 1.

If the value of right pointer match with the item then assign right pointer to left pointer. Assign data list length to right pointer and next element in item list to item, go to step 3.Repeat this process till all the element in the item list will be processed.

IV. Algorithm : Pseudo Code

"arr" list of data element and "Data" is the list of value to be search.

```
Multi Values Bidirectional Square Root Search
(arr,Data)
        Sqr:=square root of "len"
        Len:=length[arr]-1
        i:=0
        l:=len
        f:=0
        for f = 0; f <length[Data]; f++
        {
         item = data[f];
         j = len;
         temp = i;
         sqr = square root of "l"
         while arr[i] != item && arr[j] != item && sqr != 1
          {
           if arr[i] < item && arr[j] > item
             {
              i = i + sqr
              j = j - sqr
             }
           else
             {
              if arr[i]>item
                {
                 j = i - 1
                 i = i - sqr + 1
                 sqr = square root of "sqr"
                }
              else
                {
                 i = j + 1
                 j = j + sqr - 1
                 sqr = square root of "sqr"
                }

             }
          }
         if arr[i] == item
           {
            Found Item at "i"
            while arr[i] == item && i!=temp
              {
               i = i - 1
              }
            if i==temp
              {
               i = i + 1
              }
            else
              i = i + 2
           }
          else
```


http://sites.google.com/site/ijcsis/
ISSN 1947-5500

```
        {
          if arr[j] == item
            {
              Found Item at "j"
              while arr[j]==item
                {
                  j = j - 1
                }
              i = j + 2

            }
          else
                {
                  if arr[i+1] == item
                  {
                  Found Item at "i+1"
                        i=i+2
                      }
                        else
                  Not Found
                  }
            }
      l = len – i
    }
```

V. Running Cost Analysis

A. Best Case:

The main structure of the algorithm depicts that there is an outer main loop within which there lies another inner loop , after number of inner while loop iterations, we will have found our item or concluded that it was not found. Multiple Values Bidirectional Square Root Search Best Case Occur when items are 1st or last element of the Data List. Let T (n) be the running time of Multiple Values Bidirectional Square Root Search on an input list of size n. then running time of Multiple Values Bidirectional Square Root Search Can be calculated as

Length of input array: n
Length of items array: r
Outer Loop runs: r
Inner Loop Runs: 1

so the running time of Multiple Values bidirectional square root search can be written as,

$$T(n) = O(r * 1)$$

B. Average and Worst Case:

The main structure of the algorithm depicts that there is an outer main loop within which there lies another inner loop , after number of inner while loop iterations, we will have found our item or concluded that it was not found. Let T(n) be the running time of Multiple Values Bidirectional Square Root Search on an input list of size n. then running time of Multiple Values Bidirectional Square Root Search Can be calculated as

Length of input array: n
Length of items array: r
Outer Loop runs: r
Inner Loop Runs: $n^{1/2} + n^{1/4} + n^{1/8} \ldots\ldots\ldots\ldots\ldots\ldots + n^{1/k}$
Where $n^{1/k} \geq 1$
Comparison Statements: c

So the running time of Multiple Values bidirectional square root search can be written as,

$$T(n) = r*(n^{1/2}(c+c+c) + n^{1/4}(c+c+c) + n^{1/8}(c+c+c) + \ldots\ldots + n^{1/k}(c+c+c))$$

$$T(n) = r*(n^{1/2}(3c) + n^{1/4}(3c) + n^{1/8}(3c) + \ldots\ldots\ldots\ldots\ldots\ldots + n^{1/k}(3c))$$

Ignoring Constants and smaller values we can write it as

$$T(n) = O(r\sqrt{n})$$

VI. Comparison Of Multiple Values Bidirectional Square Root Search With Linear Search and Binary Search

Table 1 shows the comparison of the proposed algorithm (Multiple Values Bidirectional Square Root Search) with the traditional searching algorithms on the basis of no. of iterations. Graphical view of the same analysis is presented in the Figure 1.

Multiple Values Bidirectional Square Root Search															
Index	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
Data	**5**	**12**	**14**	**26**	**33**	**37**	**42**	**48**	**57**	**59**	**69**	**72**	**85**	**88**	**91**
For 57															
1st iteration									**NF**						
2nd iteration									**NF**						
3rd iteration									**F**						
For 85															
4th iteration													**NF**		
5th iteration													**F**		
For 91															
6th iteration													**F**		
Binary Search															
Index	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
Data	**5**	**12**	**14**	**26**	**33**	**37**	**42**	**48**	**57**	**59**	**69**	**72**	**85**	**88**	**91**
For 57															
1st iteration									**NF**						
2nd iteration									**NF**						
3rd iteration									**NF**						
4th iteration									**F**						
For 85															
5th iteration													**NF**		
6th iteration													**NF**		
7th iteration													**NF**		
8th iteration													**F**		
For 91															
9th iteration															**NF**
10th iteration															**NF**
11th iteration															**NF**
12th iteration															**F**
Linear Search															
Index	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
Data	**5**	**12**	**14**	**26**	**33**	**37**	**42**	**48**	**57**	**59**	**69**	**72**	**85**	**88**	**91**
For 57															
1st iteration									**NF**						
2nd iteration									**NF**						
……..									**NF**						
……..									**NF**						
9th iteration									**F**						
For 85															
10th iteration													**NF**		
……..													**NF**		
……..													**NF**		
22nd iteration													**F**		
For 91															
23rd iteration															**NF**
……..															**NF**
……..															**NF**
37th iteration															**F**

Table 1: Multiple Values Bidirectional Square Root Search Vs Binary Search and Linear Search

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

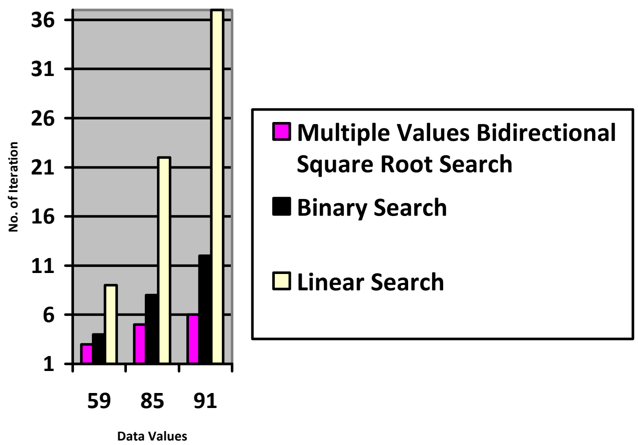


Figure 1: Multiple Values Bidirectional Square Root Search v/s Binary Search and Linear Search

It can be observed from the Table 1 and graph presented in Figure 1 that performance of proposed algorithm (Multiple Values Bidirectional Square Root Search) is far better than the linear search. Also it shows good result in comparison with Binary Search and show good competition.

Table 2 shows the comparison of the proposed algorithm (Multiple Values Bidirectional Square Root Search) with the traditional searching algorithms on the basis of No. of comparisons. Graphical view of the same analysis is presented in the Figure 2. Total values in data list = 100000(1…..100000)

Values	Comparisons		
	Linear Search	Multiple Values Bidirectional Square Root Search	Binary Search
1	1	1	48
948	948	9	40
10112	10112	32	43
30336	30336	67	38
50560	50560	67	32
69032	69032	75	38
89888	89888	33	37
97156	97156	10	41
100000	100000	1	35

Table 2: Multiple Values Bidirectional Square Root Search Vs Binary Search and Linear Search

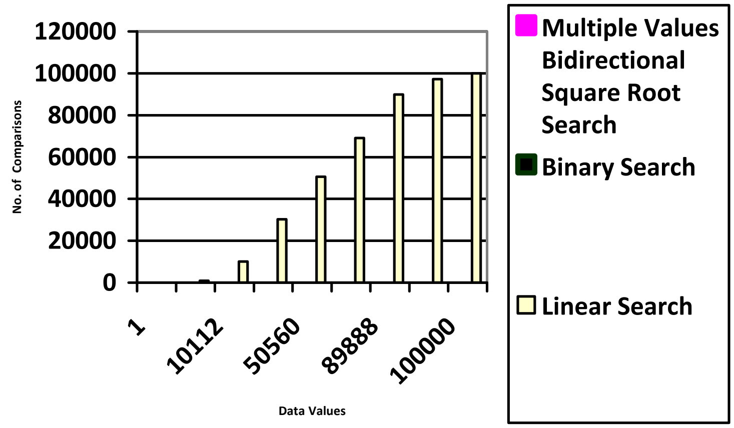


FIGURE 2: MULTIPLE VALUES BIDIRECTIONAL SQUARE ROOT SEARCH V/S BINARY SEARCH AND LINEAR SEARCH

It can be observed from the Table 2 and graph presented in Figure 2 that performance of proposed algorithm (Multiple Values Bidirectional Square Root Search) is far better than the linear search. Also it shows good result in comparison with Binary Search and show good competition.

Table 3 shows the comparison of the proposed algorithm (Multiple Values Bidirectional Square Root Search) with the traditional searching algorithms on the basis of execution time (in ms). Graphical view of the same analysis is presented in the Figure 3. Total values in data list = 1000000(1…..1000000)

No. of Elements	Execution Time(millisecond)		
	Linear Search	Multiple Values Bidirectional Square Root Search	Binary Search
50000	19578	15.6	15.6
100000	40950	31.2	46.8
200000	80979.6	62.4	93.6
300000	122226	78	109.2
500000	210678	109.2	187
600000	350107	124	234

Table 3: Multiple Values Bidirectional Square Root Search Vs Binary Search and Linear Search

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

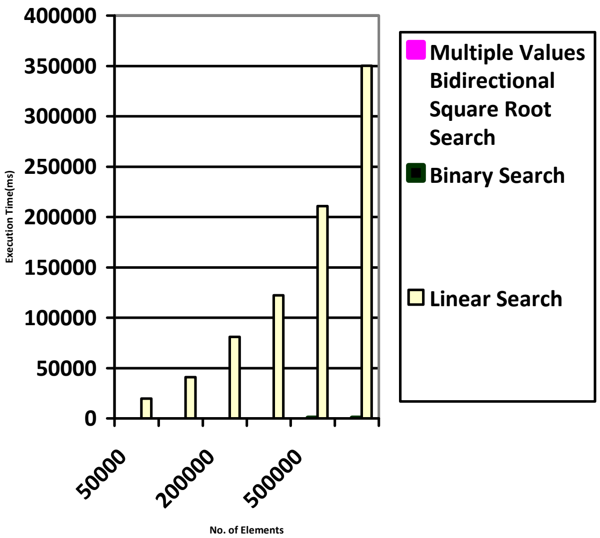


Figure 3: Multiple Values Bidirectional Square Root Search v/s Binary Search and Linear Search

It can be observed from the Table 3 and graph presented in Figure 3 that performance of proposed algorithm (Multiple Values Bidirectional Square Root Search) is far better than the linear search. Also it shows good result in comparison with Binary Search and show good competition.

VII. CONCLUSION

By analyzing the graphs above, it can be easily examined that Multiple Values Bidirectional square root search efficient then linear search and also give good competition to binary search. The common thing between Multiple Values Bidirectional Square Root Search and Binary Search is that for both, the list of element should be sorted. In future, we are foreseeing to come up with a new searching technique, which hopefully will be more efficient.

ACKNOWLEDGMENT

We acknowledge the support and contribution from our loving and caring families, teachers and friends in continuing the education and research. With their moral and financial support, we are in higher education, which encourages us to write this research for the ease of human kind. Thank you all.

REFERENCES

[1] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. Introduction to Algorithms, Second Edition. MIT Press and McGraw-Hill, 2001. ISBN 0-262-03293-7. Section 1.1: Algorithms, pp.5–6.
[2] D. A. Bailey, Java Structure:Data Structure in Java for Principled Programmer, 2nd ed. McGraw-Hill, 2003.
[3] R. Sedgewick, Algorithms in C++. Reading, Massachusetts: Addison-Wesley, 1992.
[4] Donald Knuth. The Art of ComputerProgramming, Volume 3: Sorting and Searching, Third Edition. Addison-Wesley, 1997. ISBN 0-201-89685-0. pp. 106–110 of section 5.2.2: Sorting by Exchanging.
[5] Binary_search, (2010), Binary search, http://en.wikipedia.org/wiki/Binary_search, Accessed January 01, 2010
[6] Donald Knuth. The Art of Computer Programming, Volume 3: Sorting and Searching, Third Edition. Addison-Wesley, 1997. ISBN 0-201-89685-0., pp. 138–141, of Section 5.2.3: Sorting by Selection.
[7] Seymour Lipschutz. Theory and Problems of Data Structures,Schaum's Outline Series: International Edition, McGraw- Hill, 1986. ISBN 0-07-099130-8., pp. 324–325, of Section 9.4: Selection Sort.
[8] Seymour Lipschutz. Theory and Problems of Data Structures,Schaum's Outline Series: International Edition, McGraw- Hill, 1986. ISBN 0-07-099130-8., pp. 322–323, of Section 9.3: Insertion Sort.
[9] Linear_search, (2010), Linear search, http://en.wikipedia.org/wiki/Linear_search, Accessed January 01,2010.

Chunk Sort

Syed Zaki Hassan Kazmi
Department of Computer Science
IQRA University Islamabad Campus
Pakistan
zaki.mzd@gmail.com

Syeda Shehla Kazmi
Department of Computing & Mathematics
Manchester Metropolitan University,
United Kingdom
shehla_kazmi@hotmail.com

Syeda Sobia Hassan Kazmi
Department of Computer Science
The University Of Azad Jammu And Kashmir
Muzaffarabad A.K, Pakistan
fajar_zara@hotmail.com

Syed Raza Hussain Bukhari
Department of Computer Science
The University Of Azad Jammu And Kashmir
Muzaffarabad A.K, Pakistan
razabukhari@hotmail.com

***Abstract*—the objective of this paper is to develop new efficient sorting technique known as Chunk Sort. Chunk sort is based on the idea of Merge Sort. In this sort divide main data list into number of sub list then sort sub list and combine them until the whole list become sorted. The difference between merge sort and chunk sort is, merge sort merge two sub list in to one, but chunk sort merge three sub list in to single sorted list. It is fast then many existing sorting techniques including merge sort and quick sort. Running Cost analysis and the result obtained after implementation are provided in the graphical form with an objective to compare the efficiency of the technique with quick sort and merge sort, as both are consider as efficient and fast one.**

Keywords-Sorting;Merge Sort; Quick Sort.

I. INTRODUCTION

Algorithms have a vital and key role in solving the computational problems, informally an algorithm is a well defined computational procedure that takes input and produces output. Algorithm is a tool or a sequence of steps to solve the computational problems [1].

An Algorithm is a step by step process to solve a particular problem in a finite amount of time and has a fundamental and key role in solving the computational problems. Informally an algorithm is a well-defined computational procedure that takes input and produces output. So a Sorting Algorithm is the step by step process by which data is organized into a particular increasing or decreasing order, i.e. in ascending and descending order. Data can be in numerical or character form in particular and can be in other forms as well. There are a lot of sorting techniques, currently used in industry and academia, to arrange the data of various forms and from different areas. Sorting is of substantial importance as the human is obsessed in keeping the ordered information/knowledge. To search the particular data efficiently the arrangement of data is very important. To facilitate the human, computers consume a substantial time in ordering the data. The computational problems always have a cumbersome effect on the researchers on one hand and open the opportunities for them on the other hand.

The study in hand proposes a new sorting technique that is tested and analyzed against merge sort and quick sort to provide its efficiency.

This paper is organized as follows; Section II presents a brief review of existing sorting algorithms. Section III presents the description of proposed solution. Section IV presents the proposed algorithm. Section V presents Running Cost Analysis. Section VI present comparison between chunk sort with merge sort and quick sort. Section VII ends with concluding remarks.

II. A BRIEF REVIEW OF EXISTING SORTING ALGORITHMS

A number of sorting techniques are currently used in the field of computer science. This section will briefly discuss some of the trendy sorting techniques among them. These are following:

A. Merge Sort

Merge sort is an O (n log n) sorting algorithm. It belongs to the family "Sorting by Merging". It is an example of the divide and conquer algorithmic paradigm. It was proposed by John von Neumann in 1945 (Cormen T. H, Leiserson C. E., Rivest R. L. and Stein C. [1990] 2001). Conceptually, a merge sort works as follows:

1. If the list is of length 0 or 1, then it is already sorted. Otherwise:
2. Divide the unsorted list into two sublists of about half the size.
3. Sort each sublist recursively by re-applying merge sort.
4. Merge the two sublists back into one sorted list.

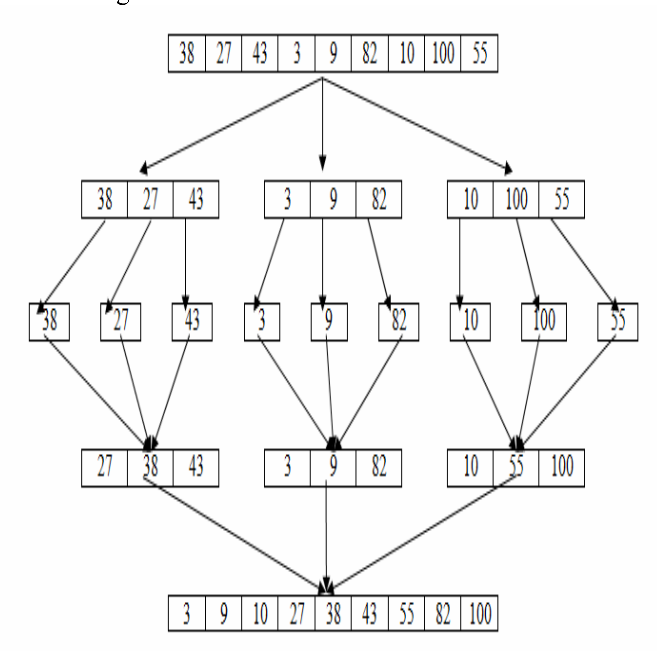
B. Quick Sort

Quick sort is another well-known sorting algorithm and base on divide-and-conquer paradigm. Its worst-case running time is $\Theta(n^2)$ having a list of n items. In spite of slow worst-case running time, quick sort is often the best practical choice for sorting the lists because it is extremely efficient on the average running time i.e. $\Theta(n\log n)$. In most real-world data it is possible to make design choices which minimize the probability of requiring quadratic time (Cormen T. H, Leiserson C. E., Rivest R. L. and Stein C. Quicksort 2001).

III. CHUNK SORT

Chunk sort is based on the idea of Merge Sort. In this sort divide main data list into number of sub list then sort sub list and combine them until the whole list become sorted. The difference between merge sort and chunk sort is, merge sort merge two sub list in to one, but in chunk sort merge three sub list in to single sorted list. It is fast then many existing sorting techniques including merge sort and quick sort. Conceptually, a Chunk sort works as follows:

1. If the list is of length 0 or 1, then it is already sorted. Otherwise:
2. Divide the unsorted list into three sublists.
3. Sort each sublist.
4. Merge the three sublists back into one sorted list.



IV. ALGORITHM : PSEUDO CODE

```
Square Root Sort (arr)
        temp = temperary Array
        l = length[arr]
        length = l
        t = l / 4
        sqr = 3
sqrr = 1
p2 = 0
while sqrr != l
{
   p1 = 0
   p3 = 0
   while p3 < l
   {
      r = 2
      p3 = p3 + sqr
      if p3 > l
      {
         p3 = l
      }
      while p2 != p3 - 1
      {
         p4 = p2 + 1
         p2 = p2 + sqrr
         if p2 >= p3
         {
            p2 = p3 - 1
         }
         if r % 2 == 0
         {
            if arr[p1] >= arr[p2]
            {
               r = r + 1
               j = p1
               for k = p4; k <= p2; k++
               {
                  temp[j] = arr[k]
                  j++
}
               x = p1
               while x != p4
               {
                  temp[j] = arr[x]
                  x++
                  j++
}
            }
            else
            {
             If arr[p4]<arr[p4-1]
               {
                  r = r + 1
                  j=p1
                  x=p4
                  for k = p1; k <= p2; k++
                  {
                     if arr[j] <= arr[x]
                     {
                        temp[k] = arr[j]
                        j++
                        if j == p4
                        {
                           k++
                           while k <= p2
                           {
                              temp[k] = arr[x]
                              x++
                              k++
```


http://sites.google.com/site/ijcsis/
ISSN 1947-5500

```
                    }  }}
                else
                {
                    temp[k] = arr[x]
                    x++
                    if(x>p2)
                    {
                        k++
                        while k<=p2
                        {

                            temp[k] = arr[j]
                            j++
                            k++
                        } }}}}}}

else
    {
        if temp[p1] >= arr[p2]
        {
            r = r + 1;
            j = p1
            for k = p4; k <= p2; k++
            {
                arr[j] = arr[k]
                j++
            }
            x = p1
            while x != p4
            {
                arr[j] = temp[x]
                x++
                j++

            }
        }
        else
        {

            if arr[p4] < temp[p4 - 1]
            {
                r = r + 1
                j = p1
                x = p4
                for k = p1; k <= p2; k++
                {
                    if temp[j] <= arr[x]
                    {
                        arr[k] = temp[j]
                        j++
                        if j == p4
                        {
                            k++
                            while k <= p2
                            {
                                arr[k] = arr[x]
                                x++
                                k++
                            }
                        }
                    }
                    else
                    {
                        arr[k] = arr[x]
                        x++
                        if x > p2
                        {
                            k++
                            while k <= p2
                            {

                                arr[k] = temp[j]
                                j++
                                k++
                            }

                        }

                    }

                }

            }

            else
            {
                r = r + 1
                for k = p1; k <p4; k++
                {
                    arr[k] = temp[k]

                }

            }
        }

    }

    }
    if r % 2 != 0 && p1 != p2
    {
        for j = p1; j <= p2; j++
        {

            arr[j] = temp[j]

        }
    }
    p1 = p2 + 1
    p2 = p1 + sqrr - 1

}
sqrr = sqr
p2 = sqrr - 1

if sqr >= t
{
    sqr = l
}
else
{
    sqr = sqr * 3
}
}
```


http://sites.google.com/site/ijcsis/
ISSN 1947-5500

V. RUNNING COST ANALYSIS

In order to get a better handle of what the resulting complexity might be, suppose that we denoted by T(n) the amount of time that Chunk Sort uses on an array of size n. Recall that executing a sequence of instructions will cause us to add the running time. Hence, the running time will obey the following equation:

$T(n) = T(n/3)+ T(n/3)+ T(n/3)+T_{merge}(n) = 3T(n/3)+cn$

where c is a constant, reflecting the actual number of basic operations (comparisons, tests, arithmetic operations, assignments) in the merge routine. we would like to get a closed formula for T, or at least figure out what its fastest-growing term is (so that we can figure out O() for the algorithm).To get a better grip on the problem, let us unwind T for a couple more steps:

$T(n) = 3T(n/3)+cn$

$T(n) = 3(3T(n/9)+cn/3)+cn=3^2\ T(n/9)+2cn$

$T(n) = 3^2\ (3T(n/27)+cn/9)+cn=3^2\ T(n/27)+3cn$

How many times can we continue this expansion? Until we get to T(1) which is 1(this corresponds to the base case of running on an array of size 1). Since in each step we divide n, we will reach T(1) is log3 n steps. At that point, we will have:

$T(n) = 2\log_3 nT(1)+cn \log_3\ n$

The first term above is O(n), the second is $O(n \log_3 n)$, so the whole algorithm is

$$T\ (n) = O(n \log_3 n)$$

VI. COMPARISON OF CHUNK SORT WITH MERGE SORT AND QUICK SORT

Table 1 show the comparison of the proposed algorithm (Chunk Sort) with the traditional sorting algorithms on the basis of execution time when input data list is in descending order. Graphical view of the same analysis is presented in the Figure 1.

	10000	100000	500000	1000000	2000000	3000000	4000000	5000000
Chunk Sort	7	43	226	282	505	783	1093	1397
Merge Sort	11	108	325	579	1266	2128	2707	3259
Quick Sort	107	631	3174	6677	13897	22879	29807	40652

Table 1: Chunk Sort v/s Merge Sort and Quick Sort

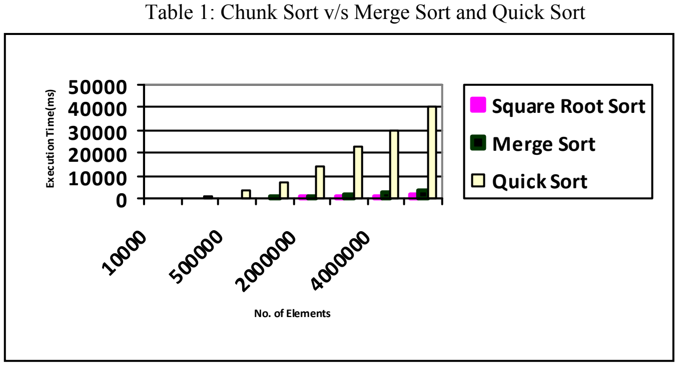


Figure 1: Chunk Sort v/s Merge Sort and Quick Sort

It can be observed from the Table 1 and graph presented in Figure 1 that performance of proposed algorithm (Chunk Sort) is far better than the Quick Sort and Merge Sort.

Table 2 shows the comparison of the proposed algorithm (Chunk Sort) with the traditional sorting algorithms on the basis of execution time when input data list is in ascending order. Graphical view of the same analysis is presented in the Figure 2

	10000	100000	500000	1000000	2000000	3000000	4000000	5000000
Chunk Sort	4	9	30	37	123	133	139	169
Merge Sort	11	109	393	608	1241	2160	2774	3584
Quick Sort	119	566	3114	6814	13838	21619	29946	40641

Table 2: Chunk Sort v/s Merge Sort and Quick Sort

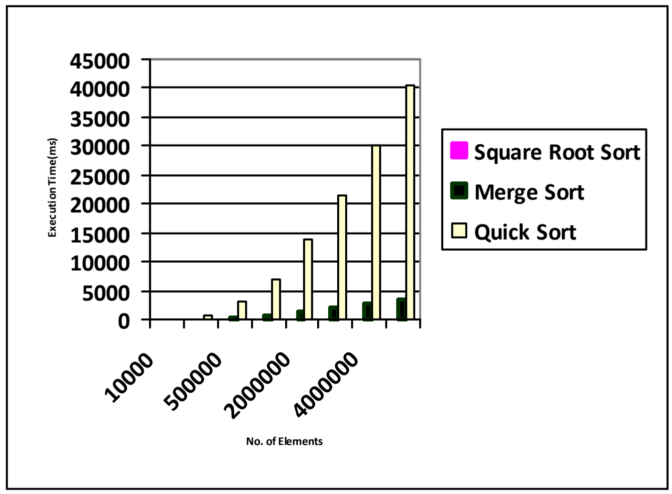


Figure 2: Chunk Sort v/s Merge Sort and Quick Sort

It can be observed from the Table 2 and graph presented in Figure 2 that performance of proposed algorithm (Chunk Sort) is far better than the Quick Sort and Merge Sort.

Table 3 shows the comparison of the proposed algorithm (Chunk Sort) with the traditional sorting algorithms on the basis of execution time when input data list is generated by random numbers. Graphical view of the same analysis is presented in the Figure 3.

	10000	100000	500000	1000000	2000000	3000000	4000000	5000000
Chunk Sort	11	129	324	546	1119	1703	2478	3105
Merge Sort	13	148	379	793	1846	2876	3608	4631
Quick Sort	173	1141	6652	13317	29559	43284	55943	75063

Table 3: Chunk Sort v/s Merge Sort and Quick Sort

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

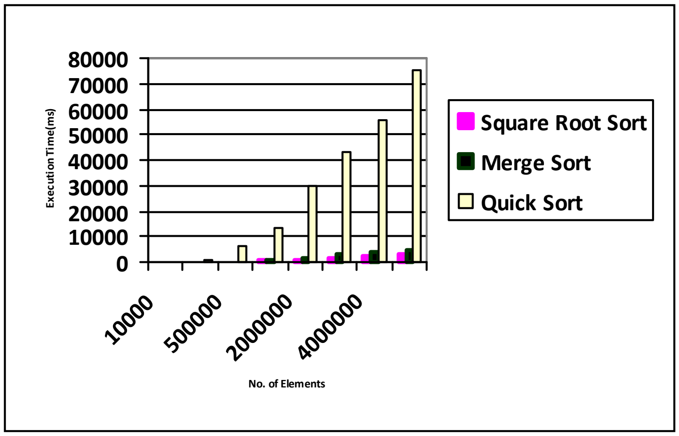


Figure 3: Chunk Sort v/s Merge Sort and Quick Sort

It can be observed from the Table 6.8 and graph presented in Figure 6.11 that performance of proposed algorithm (Chunk Sort) is far better than the Quick Sort and Merge Sort.

VII. Conclusion

By analyzing the graphs above, it can be easily examined that Chunk Sort is efficient then merge sort and quick sort. In future, we are foreseeing to come up with a new sorting technique, which hopefully will be more efficient.

Acknowledgment

We acknowledge the support and contribution from our loving and caring families, teachers and friends in continuing the education and research. With their moral and financial support, we are in higher education, which encourages us to write this research for the ease of human kind. Thank you all.

References

[1] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. Introduction to Algorithms, Second Edition. MIT Press and McGraw-Hill, 2001. ISBN 0-262-03293-7. Section 1.1: Algorithms, pp.5–6.

[2] Knuth D. (1997) *"The Art of Computer Programming, Volume 3: Sorting and Searching'',* Third Edition. Addison-Wesley, 1997. ISBN 0-201-89685-0.

[3] Philippas T., Yi Z. (2003), *"A Simple, Fast Parallel Implementation of Quicksort and its Performance, Evaluation on SUN Enterprise 10000".* IEEE- Euro micro Conference on Parallel, Distributed and Network-Based Processing (Euro-PDP'03). ISBN: 0-7695-1875-3/03

[4] Iqbal Z., Gull H. and Muzaffar A. W. (2009) "*A New Friends Sort Algorithm*". 2nd IEEE International Conference on Software Engineering and Information Technology, ISBN 978-1-4244-4520-2, pp 326-329.

[5] Merritt S. M. (1985), "*An inverted taxonomy of Sorting Algorithms*". ACM Journal, Programming Techniques and Data Structures, Vol. 28, Number 1.

[6] Green, C., Barstow. D. (1978), "*On program synthesis knowledge*". Artif. Infd. 10, pp. 241-279.

[7] Hildebrandt P. and Isbitz H. (1959) *"Radix Exchange—An Internal Sorting Method for Digital Computers".* Journal of the ACM Volume 6, Issue 2, ISSN: 0004-5411 pp: 156 – 163.

[8] Cormen T. H, Leiserson C. E., Rivest R. L. and Stein C. [1990] (2001). "*Introduction to Algorithms*", 2nd edition, MIT Press and McGraw-Hill, ISBN 0-262-03293-7, pp. 27–37. Section 2.3: Designing algorithms.

[9] Cormen T. H, Leiserson C. E., Rivest R. L. and Stein C. (2001). "*Introduction to Algorithms*", 2nd edition, MIT Press and McGraw-Hill, ISBN 0-262-03293-7, pp. 145-149. Section 7.1: Quick Sort.

[10] Cook C. R., Kim D. J. (1980) "*Best Sorting Algorithms for Nearly Sorted Lists*", Communication of ACM, Vol. 11, Issue 11, pp 620-624.

[11]Wainwright R.L. (1985) "*A Class of Sorting Algorithms based on Quick Sort*", Communication of ACM, Vol. 28, Issue 4, pp 396-402.

Top-Down Approach for the Development of Scheduling Mechanism in Packet-Switching Networks

Yaser Miaji and Suhaidi Hassan

InterNetWorks Research Group, UUM College of Arts and Sciences
Universiti Utara Malaysia, 06010 UUM Sintok, MALAYSIA
ymiaji@internetworks.my suhaidi@ieee.org

Abstract
Resource sharing in network society is common particularly with the current enormous increase in the number of Internet users with respect to limited resources. The emergence of new application such as real-time applications with more sensitivity and odd behavior required fundamental change in resource sharing policy. The delay which caused by congestion could sensationally demolish all the aspiration for delivering such sensitive application with respect to its restrictions. Scheduling mechanism plays an essential and crucial part of this function since queuing delay is the largest delay which experienced by the packet. Therefore, this paper is looking at the evolution of scheduling mechanism in the academic environment. We present the development in top-down approach form the last proposal to the very beginning. Although such approach provides a comprehensive knowledge, intensive information and rigorous tracking of the evolution of scheduling mechanism, the results shows that there is no much changes in the principle of the scheduling except for most recent scheduler named as Just Queuing.

Keywords: QoS, real time application, scheduling, queueing, network layer.

1. Background

The interest in licking the issue of congestion or flow control in network established from the first discovery and increase popularity of the Internet in 1967 or earlier. The congestion problem was not fully exposed at the presence of ARPANET based network due to the bandwidth informality and switching node identicalness, and the excessiveness of the capacity particularly in the department of defense in early 80's [36],[37],[38]. However, as the use of the network deployed and the popularity increase, the issue grows and the demand for an optimal or tentative solution becomes obvious. Since that time there has been an intensive effort from the scholar and researchers to solve the congestion control problem.

The problem get worse by the presence of novel traffic with different characteristics for application called real time application such as video and voice application. Another cause of this demand is the user himself. Users start to complain and request for much effective service. RFC 1193 identifies the user requirements for reliable QoS which depend on the service level of agreement. There are several attempts in different network layers to solve the issues of congestion control.

The attempts in solving the congestion problem in network layer were popular in 90's. Researchers tend to schedule the packet transmission to fulfill the requirements for different network traffic. There was a comprehensive study of queueing algorithms and scheduling disciplines to control the congestion in the network. The argument surrounded the effectiveness in term of bounding delay and packet loss rate, the efficiency in term of users' protection from misbehaved users, and the fairness of a scheduler is quite popular in the network community by the meddle of 90's. This argument result in a discovery of several scheduling mechanisms which based on two main categories namely: timestamp scheduling and round robin scheduling.

However, the effort toward the scheduling solution has been humbled because both round robin and timestamp scheduler have failed to achieve the requirements of all the scheduler's requirements; fairness, bounded delay, loss rate, and low complexity. The attempts have been changed toward another network layers. However, there are a few researchers conducted in late 1990's as well as 2000 in the scheduling area. Nevertheless, almost none of the approaches have been implemented in the market place. For a certain purpose there is a believe that this area should be revived. According to Internet coaching library, the Internet users have grown just over 342.2% since 2000 [28]. Consequently, the usage will suffer from a magnificent growth. Additionally, the emergence of new application such as online gaming and three dimensional (3D) games and application, results in a demand for distinguish effort in the area of scheduling and queuing discipline. Therefore, as been claimed by Floyd [13], there will be a negative impact of best effort traffic and unfairness of the competing traffic.

This article will demonstrate chronologically how the attempts toward timestamp based scheduling have been evolved. Furthermore, the benefit and the drawbacks of using the mechanism will be presented. Also, there will be a brief explanation of the mathematical, conceptual and

implementation issue. The key success of the scheduler in the market will be highlighted. This paper will stimulate the research thinking to identify the importance and the ability of scheduling in routers to enhance QoS for real time application. In addition it will assist the researcher to distinguish the key failure of other proposed mechanisms which have not been implemented in real routers. Next section will present the importance of scheduling mechanism in routers for supporting QoS for real time applications over other solution on different layers such as increase the bandwidth and improve the software capability. This will followed by identification of the two research directions and specifies the timestamp category to be illustrated. The following sections have been organized chronologically starting from the discovery of the congestion and end with the current evolution. The final section is summary and conclusion.

2. The Significance of Scheduling Mechanism

There are several reasons to believe in the effectiveness and the efficiency of scheduling and queueing in router to augment QoS for real time application. The use of user datagram protocol (UDP) in transport layer as the primary protocol for transporting real time application has limited the improvement of the enhancement of QoS in this particular layer. Although, the standardization of datagram congestion control protocol (DCCP) to provide sustainability for UDP, the reliability issue still available. Therefore, the scholar effort to provide enhancement for QoS for real time application is less efficient. Furthermore, it has been declared in RFC2309 that congestion avoidance mechanisms in TCP are not sufficient in all circumstances. Hence, there should be another area to enhance the QoS for real time application and provide fairness among all network traffic.

The attempt to overcome the problem by constructing user software is much efficient with the cost as a side effect. Real time protocol (RTP) header compression is one technique to reduce the real time application header which consequently reduces the delay. Increasing the bandwidth to accommodate the large amount of real time application data is not the adequate solution for the problem. With the presence of first-come-first-serve (FCFS) technique the issue will increase. The hanger for bandwidth will increase even with the potential increase of the bandwidth as the real time application files is significantly large, especially, for video streaming and conferencing. Although, video application is mush tolerant to the delay and loss then the voice, the emergence of new application such as IPTV make customer demand constrains this tolerance.

3. Study Trend

Researchers took two different paths in order to achieve the goal of solving the congestion problem in routers using scheduling. The first approach is using timestamp scheduler. The aim of this approach is to approximate the ideal process sharing which called generalized process sharing (GPS). Researchers involved in this class seek for better fairness and bound the delay to enhance the QoS for real time application. The side effect of this research is negligence of complexity issue which makes their scheduler much complicated (with the complexity of (O(n)) than to be implemented in real routers. In addition the complexity problem increase as the network organized hierarchically.

The second approach was much simpler. Researchers involved in this category were aware of the complexity issue. However, they were not conscious of the fairness and bounding delay issues. The result is simpler scheduler with fairness and bounding delay trade off. Moreover, the issue of fairness increase as the network tends to be hierarchy. Therefore, this article will be in favor of scheduling mechanism using timestamp scheduler which supports real time application characteristics. The next section will demonstrate the chronological evolution of timestamp scheduler.

4. The Slothful Age

During the last two stages, which compose of one decade, there are at least fourteen disciplines have been proposed to combat the issue of scheduling in routers using the time stamp principle. On the other hand, this stage constitutes the same period with much less proposed disciplines. Furthermore, during the last three years the effort toward improving the scheduler has experience an indolence which could be, obviously, seen from the literature.

Simple weighted fair queueing (SWFQ) proposed by [50] has adopted the concept of proportional rate server (PRS). Basically, when the backlog occurs, among the backlogged connections, only the set of connections with the minimum potential at time t is served. Each connection in this set is served with an instantaneous rate proportional to its reservation, so as to increase the potentials of the connections in this set at the same rate. The author claims that his approach has overcome the system potential complexity issue by proposing the execution of the re-calibration every time the transmission completed. However, this attempt could introduce new implementation complexity issue which could critical.

Xiaohui et al. [52] demonstrate a basic proposition of one timestamp per queue (OTPQ) which could be used in the WFQ as well as the one timestamp per packet (OTPP) approach. The article presents some mathematical approaches beside the implementation and analysis. There are some disadvantages regarding their proposition. Firstly, any failure in calculating the OTPQ will result in crucial degradation in the fairness and inaccuracy. Furthermore, the proposition was impractical.

 http://sites.google.com/site/ijcsis/
ISSN 1947-5500

New start potential fair queueing (NSPFQ) is proposed in [31] and proposed as mean starting potential fair queueing in (MSPFQ) in [32] as an enhancement for SPFQ which proposed in [44]. There a slight discrepancy between both algorithms. NSPFQ recalibrates the system virtual time using the maximum timestamp increment (MTI), which defines as a constant value that determine at the system setup and mathematically as the result of the division of maximum packet length and a rate, while it uses the system virtual time for a newly arrived packet as the last calibrated system virtual time added by the elapsed real time between two calibration events. Nevertheless, since the NSPFQ is actually, a PRF, it still has the system complexity of *O log(n).*

A new method of scheduling has proposed by [41] which called greedy fair queueing (GrFQ). The concept of GrFQ is to seek of the minimization of the maximum difference in the normalized service received by any two flows when the next packet transmission completes. Obviously, this principle has been inspired by SWFQ discipline. The author uses the relative fairness bound (RFB) to prove the fairness of the algorithm. In addition, according to the simulation result, the discipline bounding delay is approximately similar to WFQ. However, its complexity is *O(log N)* with respect to the flow which is still high.

Since there are several issues associated with the implementation of HPFQ [1], Ju [29] introduced new method which augments the mechanism and called novel HPFQ. The principle of the proposition is to divide the scheduling task in to four server; hard-QoS server scheduling, soft-QoS server scheduling, best-effort server scheduling and co-scheduling among the previous mentioned three servers. The rest of the algorithm is typical to the HPFQ. With this sort of division of tasks, it will involve practical complexity.

Lee et al. [33] presents a new scheduling mechanism which adopts the virtual clock principle and called worst-case fair weighted fair queueing with maximum rate control (WF^2Q-M). WF^2Q-M claimed to be consisted of packet shaping and scheduler which enforce the maximum rate constraints with low packet loss. However, the most obvious drawback of such algorithm is potential increase in the complexity which result of the combination of both scheduler and shaper.

Finishing potential fair queueing (FPFQ) is introduced in [25] which based on PRS model. The main contribution of this algorithm is to reduce the system potential complexity which involve in all PRS based scheduler by the combination of system potential function with the function used for the determination of the next served packet. Since this approach is quite version there are, so far, no obvious issue associated with it implementation.

To some up, the PRS scheduler model in this stage constitute most of the researcher concentration. FPFQ, GrFQ, NSPFQ, and SWFQ is the PRS based scheduler and WF^2Q-M and novel HPFQ adopted different methods.

5. DISPUTE AGE

Upon the magnificent achievement of the WFQ in convincing IETF to accept the mechanism as the basic blocks for routers in 1996, researchers establish a comprehensive discussion regarding the effectiveness, efficiency and fairness of its implementation. Study conducted by [24] concurs with the fairness and the effectiveness of WFQ in its approximation of Generalized Process Sharing. The author verifies his opinion by using a mathematical and simulation model.

However, another study conducted by [2] asserts that WFQ is not good enough in providing fairness. Therefore the author comes up with another model [3]. Worst case fair weighted fair queuing (WF^2Q) is Bennett proposed model. Wf2Q approximate GPS with high probability and difference of no more one packet. In a WF^2Q system, whilst the server chooses the next packet at time *t*, it selects only from the packets that have started receiving service in the corresponding GPS at *t*, and elects the packet among them that would finish service first in the corresponding GPS. On the other hand, the time complexity of implementing WF^2Q is high for the reason that it based on a virtual time function which is distinct with respect to the corresponding GPS system. This leads to significant computational complexity due to the need for simulating events in the GPS system.

Bennett and Zhang [1] have introduced hierarchical packet fair queueing (HPFQ) which is also called enhanced WF^2Q (WF^2Q+). Their approach is similar to WF^2Q with simpler implementation. In WF^2Q +, each flow is associated with a *weight*, such that the sum of the weights of all flows is no larger than a predefined value *W*. A flow's weight specifies how much share of the capacity of the output link a flow is entitled to receive. Note that if *W* is equal to the capacity of the link, then the weights are actual bandwidth given to each flow. By keeping track of eligible times and finishing times of flows, the packets could be scheduled according to WF^2Q +. [6],[7] simulates WF^2Q + and compare its performance with SPFQ, which will be described later. In spite of the fairness and less complexity of HPFQ, there is an issue regarding the distribution of the bandwidth in the presence of hierarchical complex network. Moreover, the model has lack of ability to serve the multimedia traffics due to less-consideration of the diversity requirement of the multimedia traffic. It, also, has the inability to accommodate the dynamic flow set and insulating the similar traffic.

Start time fair queueing (SFQ) ([21],[23] applies different method with different computational method as

well for starting and finishing time for a packet. SFQ has finish number and start number. Start number of a packet arriving into inactive connection is the current round number otherwise it is the finish number of the previous packet. Additionally, round number is set to start number of the current packet. Hence, packets scheduled in the increasing order of start number. SFQ is effective in variable bit rate (VBR) application such as the video application. Its computation method is less complex compare to WFQ and WF^2Q. Nevertheless, packet sorting complexity is an implementation issue which prevents the utilization of SFQ in real routers. Furthermore, the end-to-end delay grows proportionally with the number of session [22].

Leap forward virtual clock (LFVC) introduced in [47] has some properties with a subtle discrepancy with VC. It consists of two queues; high and low priority. Packets from the high priority queue are serviced by lowest tag first. Contrary, packets from the low priority queue must still be serviced before their tag. Furthermore, it applies the punishment policy which means that any flow exceed its throughput will be postponed. There are two obvious limitations for such discipline; the latency will be increased as the number of punished queue increase which leads to the second issue which is the accumulative packets.

Stiliadis [44] presents their mechanism with the complexity of *O(1)* for time stamp and with bounded delay and reasonable fairness which called starting potential based fair queueing (SPFQ). In this algorithm the virtual time is derived from the based potential which defines as the minimum of the starting potential of all backlogged session. Additionally, the system potential is re-calibrated to the minimum of start potential every time this minimum changes. Consequently, in term of its implementation, it requires more state information to be maintained which cause a crucial implementation issue and consequently the system complexity will increase linearly with the time.

Minimum delay self clocked fair queueing (MD-SCFQ) is another algorithm which proposed by [5]. Since MD-SCFQ uses virtual finishing time of the packet as the system virtual time, the complexity of calculating system virtual time is *O(1)*. Likewise, its fairness is optimal beside its bounded delay which is reliable. Nevertheless, the recalibration of the system virtual time which passed on weighted average virtual start time of all backlogged session introduces additional computation which results in more complexity.

To sum up this stage, the complexity issue is the dominant of this stage and therefore, none of the proposed discipline has been successful in term of its implementation in real router. Even though, STFQ, SPFQ and MD-SCFQ has less complexity than WF^2Q, WF^2Q+, LFVC, and SFQ, the potential increase of the system complexity over time obstacles its implementation.

6. The Upheaval Age

As [10] approach has been approved by IETF, researchers start a fairly long discussion and the effort toward the scheduling algorithm has been raised. As the algorithm evolved, some of the side effects of using WFQ have been illuminated and several novel and enhanced algorithms have been proposed. Self clocked fair queuing (SCFQ) proposed by [18] share some similarities with WFQ. It is believed that its first conceptual proposition was by [9]. The principle of SCFQ that every packet is tagged prior of its involvement in the queue and the packet served as its tag increase. Despite the fact that it is simpler than WFQ, it has the sacrifice of the fairness and less end-to-end bounding delay as the number of session grows which makes it less concurrent.

There are some other researchers in favor of WFQ and some other disciplines oppose it. Golestani et al. [19] introduces a new algorithm, which work cooperatively with first-in-first-out (FIFO), to solve the problem of standalone FIFO in the degradation in controlling the delay and the congestion of a network. His conceptual and mathematical approach, which has been presented in several articles [15], [16], [17], [18], [19], and [20], tended to address the issue in the frame level by introducing the departing and arriving frame. This attempt considered as the basic blocks for timestamp scheduling even before the proposition of [10] as it has been mentioned in his thesis which approved in 1979 [16].

Golestian [16] claims that his approach designed for connection-oriented network. Nevertheless, there are two parts of his strategy; the packet admission and the stop-and-go. Since stop-and-go primarily tends to be implemented in the switches, it could be adopted for connectionless-oriented network. Essentially, his concept is to assign a departing and arriving frame for each packet and divide the time in to frames. The served packet should be departed in the previous time frame in order to be sent. However, the lack of work-conserving discipline and unfairness, despite the fact that his approach is better in regard of bounding delay-variance, led to less popularity of the algorithm in the market environment.

L. Zhang [55] demonstrates his new discipline which called virtual clock (VC). Primarily, he introduces his algorithm in his thesis in 1989 and his first paper in the discipline in 1991. The concept of his mechanism is to emulate the time division multiplexing (TDM) by allocating a virtual transmission time for each packet. However, its insufficient ability to provide fairness, which considered as a primary condition for designing a scheduler, led to its supersession.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

There are two disciplines which cognitively correlated to earliest deadline first scheduling (EDF); delay earliest-due-date (Delay-EDD) and jitter earliest-due-date (Jitter-EDD). However, there is a subtle difference EDF, Delay-EDD and Jitter-EDD. EDF does not provide protection to the host from misbehavior host. Delay-EDD, which proposed by [11],[12] attempts to overcome this issue by assigning a rate to each flow and compute a deadline based on packet arrival time and allocate separate rate and delay. However, its lack of providing proper bounding for delay variation (jitter), results in its failure to enhance the QoS for real time application. Jitter-EDD, which proposed by [49] achieves a satisfactory level of bounding delay variation and requires less buffer size. Basically, it maintains a head time which stored in the packet header and different from deadline and arriving time. This packet head time delays the packet to allow reconstruction and hence avoidance of jitter. Nevertheless, it utilizes the concept of non-conserving scheduler which is not preferred in term of providing QoS for real time application.

The innovation in the scheduling was not limited in the above mentioned discipline in this period, rather it has been expanded but these are the distinguished mechanisms. It titled as the revolution period because there are several novel ideas has been proposed on this period. There is diversity in implementing each mechanism either in mathematical or simulation approach. To conclude these mechanisms will be counted and grouped in two main categories; work conserving and non-work conserving discipline. The work conserving algorithms are WFQ, SCFQ, Virtual Clock and Delay-EDD. The non-work conserving algorithms are Stop-and-Go and Jitter-EDD. Finally, [54] is a distinguish paper which compares those algorithms in their underlying algorithm, performance guarantee provided, buffer space requirement, associated admission control policy and implementation issues.

7. The Breakthrough

The issue of network congestion was first successfully addressed by [36], [37], [38] and published in RFC 896. In Nagle [36], there was a clear explanation of the congestion collapse. The issue of the integrated network has been mentioned in Nagle article by introducing a solution for the problem of small-packet. This means that as different traffic with different characteristics emerge to the network, the network will suffer from providing reliability and quality for serving this new traffic. Consequently, the network should behave differently. The inhibition of sending new packets till the arrival of acknowledgement (ACK) was his basic attempt [38]. In other words, he basically eliminates the time constraint in the vendor machine. His approach was in TCP layer and quite simple and realistic but it would not serve all traffic patterns.

In addition to the small packet solution which introduces the problem of traffic diversity, Nagle proposed the gateway self defense scheme which has been later called as fair queueing (FQ). FQ considered as, in principle, the first successful approach to reduce the impact of misbehaved host and introduces the fairness demand. However, his proposal has not been implemented not even in a simulation environment. Therefore, the attention is not vital.

The famous weighted fair queueing (WFQ) was first proposed by [10]. There are several reasons for the success of WFQ. The actual demand for a technique to solve the congestion and QoS issues in the network leads to the wide acceptance of the discipline in the network community. Another key success which distinguishes this mechanism is comprehensive coverage of the mechanism description which enhanced with the simulation results. The mechanism adopts the fairness definition as a Maximization of the allocation for the most poorly treated session (maximize the minimum). Since the congestion problems dramatically grow, the Internet engineering task force (IETF) in 1996 approved WFQ as a basic building block for future integrated service network. Actually, the practical implementation of WFQ has been presented in [37] and it is primarily one of the reasons for WFQ wide deployment. Eventually, this stage has a happy ending with crucial solution, as it has been assumed, to the fairness and congestion problem.

8. The Conclusion

In this paper the importance of scheduling in router has been demonstrated. Then, the chronological evolution of timestamp scheduler has been presented. Obviously, the scheduling algorithms revolutionary developed in early and late 90's. However, in the 2000's, more attention been devoted for enhancing QoS in other layers. Most of current routers utilize the old discipline whether timestamp or round robin. Therefore, WFQ still the dominant of the scheduling mansion.

9. References

[1] J. C. R. Bennett and H. Zhang, "Hierarchical packet fair queueing algorithms," Applications, Technologies, Architectures, and Protocols for Computer Communication, pp. 143-156, 1996.

[2] J. C. R. Bennett and H. Zhang, "Why WFQ Is Not Good Enough for Integrated Services Networks," 1996.

[3] J. C. R. Bennett, H. Zhang, and F. Syst, "WF 2 Q: worst-case fair weighted fair queueing," 1996.

[4] H. B. Chiou, "Enhancing the Fairness of TCP over Internet Using an Improved Hierarchical Packet Fair Queueing Scheme," 2000.

[5] F. M. Chiussi and A. Francini, "Minimum-delay self-clocked fair queueing algorithm for packet-switched networks," 1998.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[6] N. Ciulli and S. Giordano, "Analysis and simulation of WF2Q+ based schedulers: comparisons, compliance with theoretical bounds and influence on end-to-end delay jitter," Computer Networks, vol. 37, pp. 579-599, 2001.

[7] N. Ciulli and S. Giordano, "Analysis and Simulation of WF^ 2Q+ Based Schedulers: Comparisons and Compliance with Theoretical Bounds," Lecture notes in computer science, pp. 255-272, 2001.

[8] D. Comer and M. Martynov, "Design and Analysis of Hybrid Packet Schedulers," 2008.

[9] J. R. Davin and A. T. Heybey, "A simulation study of fair queueing and policy enforcement," 1990.

[10] A. Demers, S. Keshav, and S. Shenker, "Analysis and simulation of a fair queueing algorithm," Applications, Technologies, Architectures, and Protocols for Computer Communication, pp. 1-12, 1989.

[11] D. Ferrari, "Client Requirements for Real-Time Communication Services;RFC-1193," Internet Request for Comments, 1990.

[12] D. Ferrari and D. C. Verma, "A scheme for real-time channel establishment in wide-area networks," Selected Areas in Communications, IEEE Journal on, vol. 8, pp. 368-379, 1990.

[13] S. Floyd and K. Fall, "Promoting the use of end-to-end congestion control in the Internet," IEEE/ACM Transactions on Networking (TON), vol. 7, pp. 458-472, 1999.

[14] A. Francini and F. M. Chiussi, "A weighted fair queueing scheduler with decoupled bandwidth and delay guarantees for the support of voice traffic," 2001.

[15] S. J. Golestaani, I. Massachusetts Inst Of Tech Cambridge Lab For, and S. Decision, "A Unified Theory of Flow Control and Routing in Data Communication Networks," Massachusetts Institute of Technology, Dept. of Electrical Engineering and Computer Science, 1979.

[16] S. J. Golestani, "A stop-and-go queueing framework for congestion management," ACM SIGCOMM Computer Communication Review, vol. 20, pp. 8-18, 1990.

[17] S. J. Golestani and M. Bellcore, "Duration-limited statistical multiplexing of delay-sensitive traffic in packet networks," 1991.

[18] S. J. Golestani and M. Bellcore, "A self-clocked fair queueing scheme for broadband applications," 1994.

[19] S. J. Golestani, B. C. Res, and N. J. Morristown, "Congestion-free transmission of real-time traffic in packet networks," 1990.

[20] S. J. Golestani, B. C. Res, and N. J. Morristown, "Congestion-free communication in high-speed packet networks," IEEE Transactions on Communications, vol. 39, pp. 1802-1812, 1991.

[21] P. Goyal and B. Tech, Packet scheduling algorithms for integrated services networks: University of Texas at Austin, 1997.

[22] P. Goyal and H. M. Vin, "Fair airport scheduling algorithms," 1997.

[23] P. Goyal, H. M. Vin, and H. Chen, "Start-time fair queueing: a scheduling algorithm for integrated services packet switching networks," Applications, Technologies, Architectures, and Protocols for Computer Communication, pp. 157-168, 1996.

[24] A. G. Greenberg and N. Madras, "How fair is fair queuing," Journal of the ACM (JACM), vol. 39, pp. 568-598, 1992.

[25] H. Halabian and H. Saidi, "FPFQ: A Low Complexity Fair Queueing Algorithm for Broadband Networks," 2008.

[26] D. Hogan, "Hierarchical Fair Queueing," in Basser Department of computer science, vol. Doctor of Philosophy. Sydney: University of Sydney, 1997, pp. 123.

[27] D. Hogan, "Hierarchical Fair Queuing," Technical Report 506, Basser Department of Computer Science, University of Sidney, May 1996 1997.

[28] Internet World Stats, www.internetworldstats.com-/stats.htm, 19-03-09.

[29] S. Ju and M. Najar, "A novel hierarchical packet fair scheduling model," 2003.

[30] G. B. Kramer, A. Singhal, N. K. Mukherjee, and B. Ye, "Fair queueing with service envelopes (FQSE): a cousin-fair hierarchical scheduler for subscriber access networks," Selected Areas in Communications, IEEE Journal on, vol. 22, pp. 1497-1513, 2004.

[31] D. Y. Kwak, N. S. Ko, B. Kim, and H. S. Park, "Anew Starting Potential Fair Queuing Algorithm with O (1) Virtual Time Computation Complexity," ETRI journal, vol. 25, 2003.

[32] D. Y. Kwak, N. S. Ko, and H. S. Park, "Mean starting potential fair queuing for high-speed packet networks," 2003.

[33] J. F. Lee, M. C. Chen, and Y. Sun, "WF2Q-M: Worst-case fair weighted fair queueing with maximum rate control," Computer Networks, vol. 51, pp. 1403-1420, 2007.

[34] G. Lu, "Issues and technologies for supporting multimedia communications over the Internet," Computer Communications, vol. 23, pp. 1323-1335, 2000.

[35] G. Lv and X. Zhang, "An Efficient Policy-based Packet Scheduler With Flow Cache," 2007.

[36] J. Nagle, "Congestion Control in IP/TCP Internetworks," 1984.

[37] J. Nagle, "RFC-896: Congestion Control in IP/TCP Internetworks," Request For Comments, 1984.

[38] J. Nagle, "On Packet Switches with Infinite Storage," Communications, IEEE Transactions on [legacy, pre-1988], vol. 35, pp. 435-438, 1987.

[39] A. K. Parekh and R. G. Gallager, "A generalized processor sharing approach to flow control in integrated services networks: the single-node case," IEEE/ACM Transactions on Networking (TON), vol. 1, pp. 344-357, 1993.

[40] H. Shi and H. Sethu, "An evaluation of timestamp-based packet schedulers using a novel measure of instantaneous fairness," 2003.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[41] H. Shi and H. Sethu, "Greedy fair queueing: A goal-oriented strategy for fair real-time packet scheduling," 2003.

[42] M. Song, N. Chang, and H. Shin, "A new queue discipline for various delay and jitter requirements inreal-time packet-switched networks," 2000.

[43] M. Song, J. Song, and H. Li, "Implementing a high performance scheduling discipline WF2Q+ in FPGA," 2003.

[44] D. Stiliadis and A. Varma, "Efficient fair queueing algorithms for packet-switched networks," IEEE/ACM Transactions on Networking (TON), vol. 6, pp. 175-185, 1998.

[45] I. Stoica, S. Shenker, and H. Zhang, "Core-stateless fair queueing: achieving approximately fair bandwidth allocations in high speed networks," 1998.

[46] I. Stoica, H. Zhang, and T. S. E. Ng, "A hierarchical fair service curve algorithm for link-sharing, real-time and priority services," 1997.

[47] [46] S. Suri, G. Varghese, and G. Chandranmenon, "Leap forward virtual clock: a new fair queuing scheme with guaranteed delays and throughput fairness," In Proceedings of INFOCOM&# 039; 97, 1997.

[48] P. Valente, "Exact GPS Simulation with Logarithmic Complexity, and its Application to an Optimally Fair Scheduler," 2004.

[49] A. Varma and D. Stiliadis, "Hardware implementation of fair queuing algorithms for asynchronous transfer mode networks," IEEE Communications Magazine, vol. 35, pp. 54-68, 1997.

[50] C. Wang, K. Long, X. Gong, and S. Cheng, "SWFQ: a simple weighted fair queueing scheduling algorithm for high-speed packet switched network," 2001.

[51] D. E. Wrege and J. Liebeherr, "A Near-Optimal Packet Scheduler for QoS Networks," 1997.

[52] J. Xiaohui, L. Jiandong, and G. Feng, "Two simple implementation algorithms of WFQ and their performance analysis," 2001.

[53] H. Zhang, "Service Disciplines for Guaranteed Performance Service in Packet-Switching Networks," PROCEEDINGS-IEEE, vol. 83, pp. 1374-1374, 1995.

[54] H. Zhang and S. Keshav, "Comparison of rate-based service disciplines," 1991.

[55] L. Zhang, "Virtual Clock: A New Traffic Control Algorithm for Packet Switching Networks," ACM Transactions on Computer Systems, vol. 9, pp. 101-124, 1991.

AUTHORS PROFILE

Yaser Miaji received the B.E. form Riyadh College of Technology, Saudi Arabia and M.E. degrees, from University of New South Wales, Australia. in 1997 and 2007, respectively. He is currently a doctoral researcher in Computer Science in the University Utara Malaysia. Previously, he works as a lecturer in the College of Telecommunication and Electronic in Jeddah from 1998-2206. His research interest includes digital electronics, computer network, distributed system and genetic algorithm. He is a member of InternetWorks research group, IEEE, ACM ISOC and STMPE.

Suhaidi Hassan PhD SMIEEE is an associate professor in computer systems and communication networks and the Assistant Vice Chancellor of the Universiti Utara Malaysia's College of Arts and Sciences. He received his PhD in Computing from University of Leeds in United Kingdom, MS in Information Science from University of Pittsburgh, PA and BS in Computer Science from Binghamton University, NY. He currently heads the InterNetWorks Research Group at the Universiti Utara Malaysia and chairs SIG InterNetWorks of the Internet Society Malaysia Chapter.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Survey on Text Document Clustering

M.Thangamani
Computer Technology
Kongu Engineering College
Perundurai, Tamilnadu, India
Vetha_narayana@yahoo.co.in

Dr.P.Thangaraj
Dean, School of Computer Technology and Applications
Kongu Engineering College
Perundurai, Tamilnadu, India
ctptr@yahoo.co.in

***Abstract*—Document clustering is also referred as text clustering, and its concept is merely equal to data clustering. It is hardly difficult to find the selective information from an 'N'number of series information, so that document clustering came into picture. Basically cluster means a group of similar data, document clustering means segregating the data into different groups of similar data. Clustering can be of mathematical, statistical or numerical domain. Clustering is a fundamental data analysis technique used for various applications such as biology, psychology, control and signal processing, information theory and mining technologies. For theoretical or machine learning perspective the cluster represent hidden pattern means search can be done by unsupervised learning, called data concept. For practical perspective clustering plays vital role in data mining applications such as scientific data exploration, information retrieval and text mining, spatial database applications, Web Analysis, CRM, marketing, medical diagnostics, computational, biology, cybernetics, genetics, marketing etc., in this survey we mainly concentrate on text mining and data mining. The process of extracting interesting information and knowledge from unstructured text is referred as text mining. Data mining is sorting through data to identify patterns and plot out the relationship. There are lot of algorithm based on text and data mining**.

Keywords-Text Mining, Information Retrieval and Text Mining, Spatial Database Applications, Web Analysis.

I. INTRODUCTION

Document clustering is the task of automatically organizing text document into meaning full cluster or group, such that the document in the same cluster are similar, and are dissimilar from the one in other clusters. It is one of the most important tasks in text mining. There are several number of technique launched for clustering documents since there is rapid growth in the field of internet and computational technologies, the field of text mining have a abrupt growth, so that simple document clustering to more demanding task such as production of granular taxonomies, sentiment analysis, and document summarization for the scope of devolving higher quality information from text. They involve in multiple interrelated types of objects. Co-cluster means document similarity and word similarity are defined in a reinforcing manner.

Computer network is the backbone of science and technology, merely 85% of business information is in the form of text, so logic-based programming is difficult to capture in fuzzy. In order to solve this problem we cope up with large number of words in one hand able to structures in natural language and on the other hand allow handling vagueness, uncertainty and fuzziness. Text mining is a knowledge intensive process where the user interacts with a document collection by using analysis tools. This is analogous to data mining. It extracts the useful information from data source from unstructured text. Text document used to identify simplified subset of document features that can be used to represent the particular document as the whole. This feature is said to be a representational model. Each document in a collection is made up of large number of features, so that it affects the system approach, performance and design.

The text mining system supports the presentation layer for browsing that is both dynamic and content based browsing. Text mining always use visualization tool to navigate and explore the concept pattern, this is used for expressing the complex data relationship. Text mining act as (GUI), graphical user interface, that is friendlier for interact with the graphical representation of concept pattern. The presentation layer of text mining system severs as a front end for execution of the system core discovery algorithm. This is user friendly and powerful algorithm, but very complex.

The dynamic partitioning of texts ranks top on the priority list for all business intelligence systems. However, first current text clustering approaches still suffer from major problems that greatly limit their practical applicability. Text clustering usually noticed as objective method, since it delivers one clearly defined results, which should be optimal. This is contrary, because different people need different idea in clustering the text, as the same text is viewed in terms of business as well as technical aspects. Second text clustering working in [13][12] high dimensional space, where each word is seen as a potential attribute. But clustering in high dimensional space is very difficult, since each data tends to have the same distance from all other

points. To overcome this problem we are using several techniques.[14] Third text clustering is actually useless, but if we combine with explanation of particular text were categorized into particular cluster, but this suffers from high number of feature chosen for computing cluster. Though there is lot of difficulties in high dimensional clustering, this can be eradicated by means of several algorithms our ultimate aim is to derive clustering result eventually in all space.

The remainder of this paper is organized as follows. Section II discusses some of the earlier proposed research work on text document clustering. Section III provides a fundamental idea on which the future research work focuses on. Section IV concludes the paper with fewer discussions.

II. RELATED WORK

Barry de Ville et al., [1] proposed the data-mining classification and predictive modeling algorithms that are based on bootstrapping techniques. This explains re-use of source data repeatedly that can render a holographic view of the modeled data. This holographic application is mainly used in industrial area that involves text mining warranty claims at a major international car, truck, and heavy equipment manufacturer. This paper shows, how they work, how they perform in text mining area as supplied to warranty claims. The main goal is to obtain the performances better -than -human.

Mine.T et al., [2] put forward that text mining system obtain the relationship between the topics of international conference. This paper not only says about the relation between topic and conference, but also says the relationship between information entities that users are interested.[3] arbitrary relations between concepts of a molecular biology ontology for the purpose of supporting text mining and manual ontology building. [4], [5] have given insights on work done on the WWW corpus for text mining based on ontological systems. Basically ontology is defined as specification of a conceptualization and this also refers to the subject of existence. JAVA based ontology and knowledge based framework provides a plug-and-play environment that makes it a flexible for rapid prototyping.

Qiaozhu Mei et al., says new general probabilistic model for contextual text mining that can cover several existing models as special cases. The extension of the Probabilistic Latent Semantic Analysis (PLSA) model the context variables models the context of a document. The proposed mixture model, called contextual probabilistic latent semantic analysis (CPLSA) model, can be applied to many interesting mining tasks, such as temporal text mining. PLSA [7] document act as a mixture of aspects, where each aspect is represented by a multinomial distribution. To avoid over fitting in PLSA, Blei and co-authors proposed a generative aspect model called Latent Dirichlet Allocation (LDA), which could group up the themes from document.

Miha Grcar1 et al., put fourth an approach regarding lack of software mining techniques, which means process of extracting knowledge out of source code. [8]In this paper we approach the software mining task with a combination of text mining and link analysis technique. This mainly deals with interlinks between one instance to another instance. There are mainly two approaches to build tool for software component, retrival and knowledge based approaches. First approach natural language documentation of the software components. With this approach no interpretation of the documentation is made but information is extracted via statistical analyses of the words distribution. On the other hand, the knowledge-based approach relies on pre-encoded, manually provided information Knowledge-based systems can be “smarter” than IR systems but they suffer from the scalability issue. We recently started developing an ontology-learning framework named LATINO which stands for Link-analysis and text-mining toolbox [8]. LATINO will be an open source general purpose data mining platform providing text mining, link analysis, machine learning, and data visualization capabilities.

Ingo Feinerer et al.,[9] gives a survey on text mining facilities in R and explain how typical application tasks can be carried out using our framework. We present techniques for count-based analysis methods, text clustering, text classification and string kernels [10]. Here the author introduced a new framework for text mining applications in R via the tm package. It offers functionality for managing text documents, abstracts the process of document handling and the usage of heterogeneous text formats in R. The package has integrated database backend support to minimize memory demands. An advanced metadata management is also implemented for collections of text documents to lighten the usage of large and with metadata enriched document sets, tm provides easy access to preprocessing and manipulation mechanisms such as whitespace removal, stemming, or conversion between file formats.

Alan Marwick introduced (UIMA). [10]There is lot of opportunity if we focus on using information technology to get more value from unstructured information within organizations. The new Unstructured Information Management Architecture (UIMA) framework that was recently introduced by IBM, which makes easier to develop and deploy systems that analyze unstructured media objects, like documents, in order to provide functions such as semantic search and text mining. Text mining is data mining applied to information extracted from text. How it can be combined with structured databases and data mining. This article is mainly for people who are interested in learning how the words of unstructured and structured information can be brought together.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

A.Hotho et al., [11] suggest that text clustering basically involves in high dimensional space, but it appears difficult for all types of setting. This is one of the new approaches for applying background knowledge during preprocessing in order to improve clustering results and allow for selection between results. In order to over come the difficulty, we compute multiple clustering results using k-Means. The results may be distinguished and explained by the corresponding selection of concepts in the ontology. The problem of clustering high-dimensional data sets has been researched by Agrawal et al. [12] They present a clustering algorithm called CLIQUE that identifies dense clusters in subspaces of maximum dimensionality. Hinneburg & Keim [13] show how projections improve the effectiveness and efficiency of the clustering process. Their work shows that projections are important for improving the performance of clustering algorithms.

Hossein M. Shirazi et al., projected, that in current trend, extracting information from the World Wide Web have been much familiar among all. Information extraction system defined as a system that “automatically identifies predefined set of related items" [14] , Since a lot of Web data are found in HTML pages. Since we use HTML, the extraction process requires fetching a Web document, cleaning it up using a syntactic normalization algorithm, and then, locating “objects of interest” in this Web page. This is done by first locating the minimal object-rich sub tree. Finally, the set of objects is refined to eliminate irrelevant objects. Crescenzi and colleagues [15] present a system that automatically extract data from large data-intensive Web sites their “data grabber” explores a large Web site and infers a model for it, describing it as a directed graph with nodes describing classes of structurally similar pages and arcs representing links between these pages. After pinpointing classes of interest, a library of wrappers can be generated, one per class with the help of an external wrapper generator and appropriate data can be extracted.

Embley and others [16] gave an idea to extract information automatically from HTML tables. The information is extracted in the from stepwise manner. As the first step, extract on ontology is formulated. Extraction ontology is a “conceptual model instance” that serves as a wrapper for a narrow domain of interest [16]. Second step expected attribute names and values from the ontology, third step attribute-value pairs are formed and adjusted so that they are more meaningful. In the fourth step, the extraction patterns are analyzed to refine the extracted information further. Then finally, given the input from the earlier four steps, a mapping can be inferred from the source to the target. There are several other works such as ‘road runner’,’ hidden morkov model’,’ cluster’ to study about HTML documents clustering.

Pallav Roxy, and Durga Toshniwal, implemented several approaches and that can be classified into two major categories, similarity-based approach and model-based approach. Similarity-based approach is a pair wise similarity document clustering, aiming to maximize the average similarities within clusters and minimize the average similarities between the clusters. Model-based approaches, on the other hand, attempt to learn generative models from the documents, with each model representing one particular document group. Several approaches have been so far proposed for document clustering from mid nineties. New technique such as self-organizing map [18], mixture of Gaussians [19], spherical k-mean[20], bi-secting k-means [21], mixture of multinomial [22, 23]. K-means is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The main idea is to define k centroids, for each cluster. These centroids should be placed far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest centroid, then need to re-calculate k new centroids, After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new centroid. A loop has been generated. As a result of this loop it is seen that the k centroids change their location step by step until no more changes are done, finally this algorithm aims at minimizing an objective function.

Shady Shehata put forward new view called Vector Space Model (VSM) .Vector Space Model (VSM) [24] which is a widely used for representing data for text classification and clustering. It says each document as a feature vector of the document. Each feature vector contains term-weight. Selecting and weighting these features accurately affect the result of the clustering algorithm substantially [25] [26]. Incorporating semantic features from the WorldNet [27] lexical database is one of best approaches that have been tried to improve the accuracy of text clustering techniques. In this paper he also introduced new semantic-based model. [28]The proposed model captures the semantic structure of each term within a sentence rather than the frequency. Each sentence in a document is labeled by a semantic role labeler, it can be either a word or phrase dependent on the semantic structure of the sentence.[29] Based on this analysis, each term assign some weight. The terms that have maximum weights are extracted as top terms. Synonyms of each word are added to the term vector. These concepts are analyzed on the sentence. Top terms and used in text document clustering. When a new document is introduced to the system, the proposed model can detect a concept match from this document to all the previously processed documents in the data set by scanning the new document and extracting the matching concepts.

There are several methods for exploiting correlations between terms in document clustering. The method calculates similarity between documents based on the statistical correlations between their terms and then uses

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

these similarities to group documents into clusters. For this we analyses different similarity model. In VSM, terms are independent and accordingly ignore any semantic relations between them. This implies that the proximity between documents is not similar. In addition to that, redundancy increases the dimensionality and affects the performance of algorithms. This disadvantages can be overcome by Wang et al., [31] proposed generalized VSM (GVSM) which represent the documents as vector in a non-orthogonal basis of terms. The similarity between documents is calculated based on the similarity between their terms. Here we obtain high performance

Hinneburg & Keim [13] stated that projection is one the important factor for ontology based text clustering, since projection improves effectiveness and efficiency of the clustering process. They do not mainly concentrate on clustering quality; rather it concentrates on clustering performance. There are different projection technique, each technique consider being important one. Since every individual technique have different output performance. Schuetze and Silverstein [31], these two authors had a detailed study about the different projection scheme. They aimed mainly on local and global projection. Local projection deals with 'how document maps with different subspace'. where as global projection select the relevant data for all documents using latent semantic indexing .although these projection have good performance effect, they are not suitable for real-time application, since it often does not coincide with the projection most suitable for humans to solve a particular task. So here comes the new proposal from another two authors named, Devaney and Ram describe feature selection for an unsupervised learning task, called conceptual clustering. They discuss about feature selection based on an existing COBWEB conceptual clustering system. In this evaluation they show that feature selection significantly improves. The drawback that Devaney and Ram face, however, is that COBWEB is not scalable like K-Means. Hence, for practical purposes of clustering in large document repositories, COSA (Concept Selection and Aggregation) and Term Selection (TES) seems to be suited.

COSA involves in two stages. First stage COSA maps terms and concepts using a shallow and efficient natural language processing system. Second stage, COSA uses the concept subsequent clustering. COSA comprise a tokenizer based on lexical analysis, [32]. It scans the text in order to identify boundaries of words and complex expressions like "$20.00" or "United States of America", and to expand abbreviations. Lexical analysis includes two steps, first it says about identification of the canonical common word and then recognize named entities.TEM is for preprocessing, and says feature vectors from Siver, but for only few terms, hence, it produces a low dimensional representation.

Li haiying et al.,[33] spotted about real-time clustering that involves in two steps process, the first step involves in extraction process that is partial parser and a shallow stemmer are invoked here, it can be used for both linguistic and statistical methods, in order to reduce the term variations within the returned text snippets the system introduce normalization algorithms. This avoids redundancy, for example: "games downloads" and "download games" the second step is mainly for clustering purpose, that is combining both linguistic clustering and statistical clustering. They generate hierarchical clustering for grouping of similar documents. This is done in real-time without any predefine grouping, pre-build knowledge base, or pre-processing of all the document collections used by the search engines. The clustering algorithms allow the same document to be in multiple clusters. This reflects the fact that different people usually will group same information differently.

III. Future enhancement

In future we plan to investigate various ways of constructing the keyword list and apply different clustering methods. We consider clustering the keywords to construct a new keyword space. We will also apply a stability-based criterion for determining the optimal number of clusters. In particular Major initiatives for text mining have resulted in the proposals for an Open Text Mining Interface (OTMI) and NIH's common Journal Publishing Document Type Definition (DTD) that can provide semantic cues to the machines to answer specific queries contained within the text.

IV. Conclusion

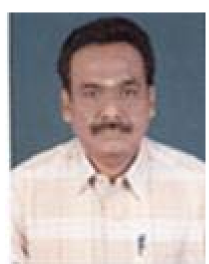
The volume of text data in the web is increasing exponentially, it makes difficult for searching purpose, several search engines in the web makes it possible to retrieve web documents by usual text database. However, users may not judge easily whether the documents have useful information, especially in the case that given keywords have wide concept, in order to .retrieve efficiently web documents, so here came the technique called "clustering". In this article, we discussed about introduction of field of text mining. Therefore, we motivated this field of research, and gave more formal definition of the terms used and presented a brief overview of currently available text mining methods, their properties and their application to specific problems. Even though, it is impossible to describe all algorithms and applications in detail, but our ideas will be interesting to every reader to provoke for their further studies. We already know that "necessity is the mother of invention", while reading this paper, most of them can have lot of questions in them. This will strive path to have a new invention in the field of text document clustering.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

REFERENCES

[1] Barry de Ville, "Text Mining with Holographic, " Decision Tree Ensembles, SAS Institute Inc., Cary, NC, 2002
[2] Unsupervised Learning of Semantic Relations for Molecular Biology Ontologies, Ciaramita, M.Gangemi, et al, "Unsupervised Learning of Semantic Relations for Molecular Biology Ontologies, " In Proceeding of the 2008 Conference on ontology Learning and Population, Bridging the Gap between Text and Knowledge,2003.
[3] P. Buitelaar and P.Cimiano, "Frontiers in Artificial Intelligence and Applications," vol. 167. IOS Press, Amsterdam, The Netherlands, 91-104.
[4] S. Pierre and Blondel. V, "Automatic Discovery of Similar Words, Survey of Text Mining, " Clustering, Classification, and Retrieval Springer (Ed.) (2008), Page 25
[5] Castellano. M, Mastronardi.G, Aprile.A, and Tarricone. G, "A Web Text Mining Flexible Architecture,"2006.
[6] T.Hofmann, " Probabilistic latent semantic analysis, " InProceedings of UAI'99.
[7] Helm.R, Maarek.Y, "Integrating Information Retrieval and Domain Specific Approaches for Browsing and Retrieval in Object-oriented Class Libraries, " In Proceedings of Object-oriented Programming Systems, Languages, and Applications, 47–61, ACM Press, New York, USA (1991).
[8] Grcar.M, Mladenic.D, Grobelnik. M, Fortuna.B, Brank.J, "D2.2: Ontology Learning Implementation," Project report IST-2004-026460 TAO, WP 2, D2.2 (2006).
[9] Karatzoglou.A, Feinerer. I (2007), "Text Clustering with String Kernels in R, " In R Decker, HJ Lenz (eds.), Advances in Data Analysis (Proceedings of the 30th Annual Conference of the Gesellschaft f ur Klassi kation e.V., Freie Universit at Berlin, March 8{10, 2006),"Studies in Classication, Data Analysis, and Knowledge Organization, pp. 91
[10] Ingo Feinerer, Kurt Hornik, David Meyer, " Text Mining Infrastructure in R, "March 2008, Volume 25, Issue 5
[11] Alan Marwick,technical lead,IBM, "Text Mining for associations using UIMA, "feb2006.
[12] R.Agrawal, J.Gehrke, D.Gunopulos, and P.Raghavan, "Automatic subspace clustering of high dimensional data for data mining applications, "In Proc. of the ACMSIGMOD Int'l Conference on Management of Data, Seattle, Washington, June 1998. ACM Press, 1998.
[13] A.Hinneburg and D.A.Keim, " Optimal grid-clustering, Towards breaking the curse of dimensionality in high-dimensional clustering, " In Proc. of VLDB-1999, Edinburgh, Scotland, September 2000. Morgan Kaufmann, 1999.
[14] P. Bradley, U. Fayyad, and C. Reina, " Scaling clustering algorithms to large databases, " In Proc. of KDD-1998, New York, NY, USA, August 1998, pages 9–15, Menlo Park, CA, USA, 1998. AAAI Press.
[15] Zamir, O.Etzioni, "Web Document Clustering, A Feasibility Demonstration, " in Proceedings of the 21st International ACM SIGIR Conference on Research and Development.
[16] D.Buttler, L.Liu, and C.Pu, "A fully automated object extraction system for the world wide web," in Proc. Int. Conf.Distrub.Comput. Syst., 2001, pp. 361–370 ormation Retrieval, ACM 1-58113-015-5 8/98, Melbourne, Australia, 1998.
[17] V. Crescenzi, G. Mecca, P. Merialdo, and P. Missier, "An automatic data grabber for large Web sites," in Proc. VLDB, 2004, pp. 1321–1324.
[18] T. Kohonen, S. Kaski, K. Lagus, J. Salojrvi, J. Honkela, V. Paatero, A. Saarela, "Self organization of a massive document collection", IEEE Trans. Neural Networks, vol. 11, 2000, pp. 574-585.
[19] J. Tantrum, A. Murua, W. Stuetzle, "Hierarchical model-based clustering of large datasets through fractionation and refractionation, " Proc. 8th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2002, pp. 183-190.
[20] I.S.Dhillon, D.S.Modha, "Concept decompositions for large sparse text data using clustering, " Machine Learning, vol. 42, 2001, pp. 143-175.
[21] M.Steinbach, G.Karypis, V. Kumar, "A comparison of document clustering techniques, " KDD Workshop on Text Mining, 2000, pp. 109-110.
[22] S.Vaithyanathan, B. Dom, "Model-based hierarchical clustering, " Proc. 16th Conf. Uncertainty in Artificial Intelligence, 2000, pp. 599–608.
[23] M. Meila, D. Heckerman, "An experimental comparison of model-based clustering methods," Machine Learning, vol. 42, 2001, pp. 9–29.
[24] G. Salton and M. J. McGill, " Introduction to Modern Information Retrieval, " McGraw-Hill, 1983.
[25] P. Mitra, C. Murthy, and S. K. Pal, "Unsupervised feature selection using feature similarity," IEEE Transactions on Pattern Analysis and Machine, vol. 24, no. 3, pp. 301–312,2002.
[26] R. Nock and F. Nielsen, "Unsupervised feature selection using feature similarity," IEEE Transactions on Pattern Analysis
[27] and Machine, vol. 28, no. 8, pp. 1223–1235, 2006.
[28] G. A. Miller, "Wordnet: a lexical database for english," Commun. ACM, vol. 38, no. 11, pp. 39–41, 1995.
[29] S. Shehata, F. Karray, and M. Kamel, "Enhancing text clustering using concept-based mining model," in Proceedings of the 6th IEEE International Conference on Data Mining (ICDM), 2006.
[30] P. Kingsbury and M. Palmer, "Propbank: the next level of treebank," in Proceedings of Treebanks and Lexical Theories,2003
[31] H.Schuetze and C.Silverstein," Projections for efficient document clustering," In Proc.of SIGIR-1997, Philadelphia, PA, July 1997,pages 74–81. Morgan Kaufmann, 1997.
[32] G. Miller. WordNet: A lexical database for english. CACM, 38(11):39–41, 1995.
[33] Li haiying,zhuang zhenquan, li bin, wan ke, "A real-time C-V clustering algorithm for web-mining, "journal of electronics, January 2002.

M. Thangamani completed her B.E.,(Electronic and Communication Engineering) from Government College of Technology, Coimbatore, India. She completed her M.E.,(Computer Science & Engineering) from Anna University , Chennai, India . Now she is doing research in the field of Clustering algorithms. Currently, she is working as a Lecturer in the School of Computer Technology and Applications, Kongu Engineering College, Tamil Nadu, India. She has published 4 articles in International journals and presented many papers in 35 National and International conferences. She has authored 15 books with reputed publishers and also guided many UG projects.



Dr. P. Thangaraj received the Bachelor of Science degree in Mathematics from Madras University in 1981 and his Master of Science degree in Mathematics from the Madras University in 1983. He completed his M. Phil degree in the year 1993 from Bharathiar University. He completed his research work on Fuzzy Metric Spaces and awarded Ph. D degree by Bharathiar University in the year 2004. He completed the post graduation in Computer Applications at IGNOU in 2005. His thesis was on "Efficient search tool for job portals". He completed his Master of Engineering degree in Computer Science in the year 2007 from Vinayaka Missions University. His thesis was on "Congestion control mechanism for wired networks". Currently he is designated as Dean of School of Computer Technology and Applications at Kongu Engineering College, Autonomous institution affiliated to Anna University. His current area of research interests are in Fuzzy based routing techniques in Ad-hoc Networks.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Simulation Analysis of Node Misbehavior in an Ad-hoc Network using NS2

Rekha Kaushik
Department of Information Technology, MANIT
Bhopal, M.P, India
rekhakaushik28@gmail.com

Dr. Jyoti Singhai
Department of Electronics and Communication Engineering
Bhopal, M.P, India
j.singhai@manit.ac.in

Abstract*— Proper operation of MANET requires mutual cooperation of participating nodes. Due to presence of selfish or malicious nodes, performance of network degrades significantly. Selfish nodes drop packets coming from its neighbor nodes to conserve its energy or push forward their own packets in the buffer queue. To prevent this misbehavior, selfish nodes need to be detected and isolated from network. This paper, detect selfish nodes which are created due to nodes conserving their energy. After their detection, performance analysis of networks has done by comparing the ideal network and the network with selfish node using NS2.

Keywords- MANET, DSR, Selfish node.

I. INTRODUCTION

A mobile ad hoc network (MANET) is a self-configuring network of mobile devices connected by wireless links, in which each device is free to move independently in any direction, and will therefore change its links to other devices frequently. Each must forward traffic unrelated to its own use, and therefore be a router. Unlike networks using dedicated nodes to support basic functions like packet forwarding, routing, and network management, in ad hoc networks these functions are carried out by all available nodes.

Applications of mobile ad hoc networks range from military tactical operations to civil rapid development such as emergency search-and-rescue missions, sensor networks, and instantaneous classroom/meeting room applications.

MANET nodes are equipped with wireless transmitters and receivers. At a given point in time, depending on the nodes' positions and their transmitter and receiver coverage patterns, transmission power levels and co-channel interference levels changes. The ad hoc topology may change with time as the nodes move or adjust their transmission and reception parameters.

In Ad hoc network, each mobile terminal has a router, and host two functions: as host, terminal need to run the applications of end-users; as a router, terminal need to run the corresponding routing protocol, participate in Packet forwarding and routing maintenance according to the routing strategies and routing table. In an Ad hoc network, the routing between nodes is through intermediate nodes because the coverage area of the wireless terminal is limited. Therefore, it is called multi-hop wireless network, self-organizing network, not fixed facility network

A. Cooperation of node

The successful operation of MANET is totally dependent on the cooperation of participating nodes in communication. The lack of a fixed infrastructure in ad hoc networks forces ad hoc hosts to rely on each other in order to maintain network stability and functionality. But sometimes nodes do not work as they are intended due to conservation of their resources such as energy, memory, and bandwidth. Such nodes are called misbehaving nodes or non cooperative nodes and are of following types:

Malicious Node:

Nodes that perform active attacks with the aim of damaging other nodes by causing network outage are considered to be malicious [1], also referred to as compromised nodes. In addition, a compromised node may use the routing protocol to advertise itself as having the shortest path to the node whose packets it wants to intercept as in the so called black hole attack.

Selfish Node:

Selfish nodes [1] work in ad hoc network for their own benefit. They simply do not forward packets (data packets and/or control packets) of other nodes to conserve their own energy, or push their own packets in front of the buffer queue.

Selfish nodes disturb the performance of ad hoc network to a great extent. When a node becomes selfish it does not cooperate in data transmission process and causes a serious affect on network performance.

In this paper, simulation analysis of node misbehavior is carried out only with selfish node. Theses nodes participate

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

correctly in routing but do not forward data packets it receives for other node; so data packets may be dropped instead of being forwarded to their destination.

B. Routing Protocols

An ad hoc routing protocol [2] is a convention, or a standard, by which node decides the way to route packets between computing devices. There are basically three types of routing protocols.

Pro-active (table-driven) routing: This type of protocols maintains fresh lists of destinations and their routes by periodically distributing routing tables throughout the network. The main disadvantages of such algorithms are respective amount of data for maintenance and slow reaction on restructuring and failures.

Reactive (on-demand) routing: This type of protocols finds a route on demand by flooding the network with Route Request packets. The main disadvantages of such algorithms are high latency time in route finding and excessive flooding can lead to network clogging.

Hybrid routing: This type of protocols combines the advantages of proactive and of reactive routing. The routing is initially established with some pro actively prospected routes and then serves the demand from additionally activated nodes through reactive flooding. The choice for one or the other method requires predetermination for typical cases. The main disadvantages of such algorithms are advantage depends on amount of nodes activated and reaction to traffic demand depends on gradient of traffic volume.

This paper uses Dynamic Source Routing protocol (DSR) [3] which is a reactive routing protocol. The DSR is a simple and efficient routing protocol designed specifically for use in multi-hop wireless ad hoc networks of mobile nodes.

DSR is an on-demand routing protocol which is based on source route approach. The protocol is composed of the two main mechanisms of "Route Discovery" and "Route Maintenance", which work together to allow nodes to discover and maintain routes to arbitrary destinations in the ad hoc network. All aspects of the protocol operate entirely on-demand, allowing the routing packet overhead of DSR to scale automatically to only that needed to react to changes in the routes currently in use.

II. RELATED WORK

Marti at [4] observed increased throughput in MANET by complementing DSR with watchdog (for detection of selfish or malicious misbehavior) and pathrater (for trust management and routing policy, every used path is rated); which enable nodes to avoid malicious nodes in their routes.

In [5] a simulation study of effects of misbehaving nodes on DSR routing protocol on the basis of countermeasures is described. This paper shows that selfish node present in the network affect the end to end delay and lead to congestion in a low density network.

In COSR (Cooperative On Demand Secure Routing Protocol) [6] Fei Wang measures node reputation and Route reputation using three parameters: contribution of node, capability of forwarding and recommendation.

Zhong et al [7] propose an incentive based system named SPRITE, in which selfish nodes are encouraged to cooperate. In this system, a node reports to the Credit Clearance Service, the messages that it has received/forwarded by uploading its receipts. Intermediate node earns credit when they forward message of others' node.

III. SIMULATION SETUP

The simulation was done using NS-2 simulator [8], provides a scalable simulation environment for wireless network systems. The simulated network consists of 20 nodes placed randomly in 670x670 areas. Each node has a transmission range of 250m and moves at a speed of 10m/s. The total sending rate of all the senders of the multi-cast group, i.e. the traffic load is 1Mbps.

Every node has initial energy set to 1000 joules. During the sending and receiving of packets, node energy gets decreases. If energy of node is less than threshold, it becomes selfish, and thus node drop all packets received from neighboring nodes.

On simulating the network with given simulation parameter and considering the energy model parameters as discuss above as shown in table 1, two selfish nodes are detected , node 2 and node 4. Table 1 lists the values of common parameters used for simulation.

Table1. Simulation Parameters

Parameter	Value
Number of Nodes	20
Routing Protocol	DSR
Packet size	512 bytes
Traffic model of sources	Constant bit rate
Mobility model	Random way point
max speed	10 m/s
Initial energy of node	1000 joules
Simulation time	25 sec

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

IV. SIMULATION RESULTS

The ideal network and the network with selfish nodes are compared on the basis of node throughput, packet delivery ratio and number of packets dropped.

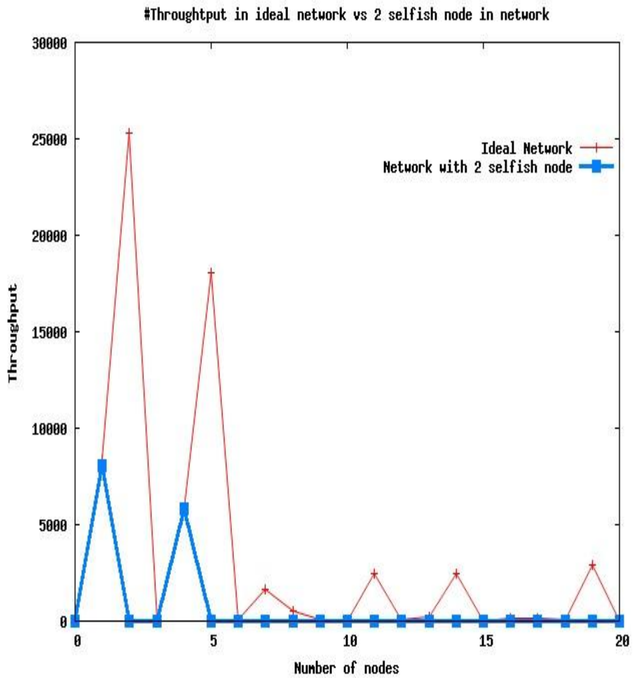


Figure 1: Throughput of ideal network vs network with selfish nodes

Node Throughput is the average rate of successful message delivery over a communication channel. This data may be delivered over a physical or logical link, or pass through a certain network node. The throughput is usually measured in bits per second (bit/s or bps).

Figure 1 shows throughput in ideal condition and throughput when there were two selfish nodes in the network.

From the figures it is shown that the throughput at node gets degraded by at node 2 and 4 as they become selfish. The overall throughput gets degraded by 80%.

Packet Delivery Ratio (PDR) is the ratio of total no. of packets sends to total no. of packets received.
Figure 2 shows the PDR when there was no selfish node in the Network and PDR when there are selfish nodes in the network..
It can be shown that the PDR increases when there are selfish nodes in the network. PDR increases by 40% when using above simulation scenario.

Number of bits dropped: From figure 3 it is seen that the number of bits dropped is more when there are a selfish node in the Network. Bit drop increases by 96% when selfish nodes are there in the network.

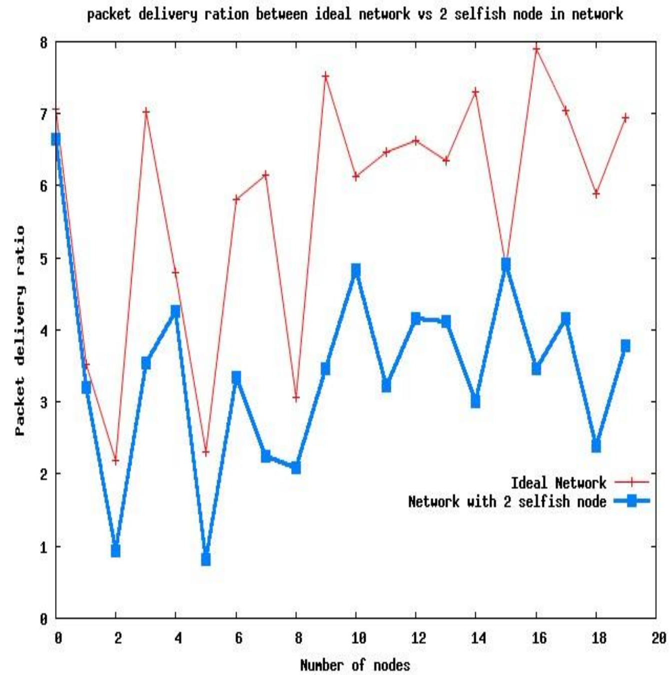


Figure 2: Packet Delivery Ratio in ideal condition vs. network with selfish nodes

+

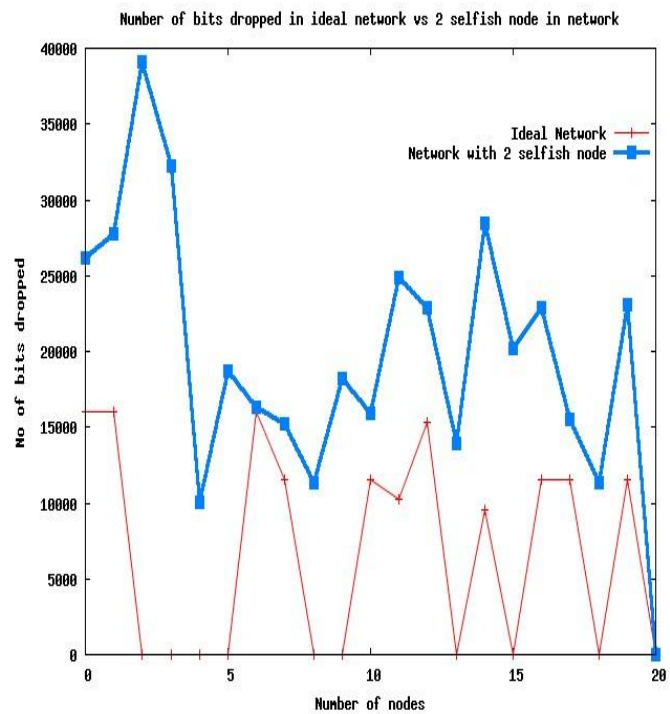


Figure 3: Number of bits dropped of ideal network vs network with selfish nodes

V. CONCLUSION AND FUTURE WORK

Misbehaving node such as selfish node in MANET can affect the performance of the overall network. This paper is the study of the selfish node impact on MANET performance. First

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

selfish nodes are detected using energy model, and then comparative analysis has been done between the networks in ideal condition i.e. the network in which there is no selfish node and the network with selfish nodes.

It has been concluded from the simulation done in NS2 that when selfish nodes are present in the network the overall network load increases on remaining nodes, hence node throughput decreases. Packet Delivery Ratio increases as nodes also forward packets with in ideal case may be forwarded by nodes which became selfish.

From the above analysis, it is concluded that either misbehaving node has been isolated from the network or some system must be include with the network which enforce cooperation among nodes to improve network performance

For future work, simulation study with malicious node is carried out and tries to get a system which motivate misbehaving node to enhance cooperation and improve network performance.

REFERENCES

[1] Matthias Hollick, jens Schmitt, Christian seipl,*''On the Effect of Node Misbehaviour in Ad hoc Network*" IEEE conference ,vol 6, pp 3759 – 3763,2004.
[2] E.Royer and C.K toh," *A Review of Current Routing Protocols for Ad hoc Networks*", IEEE Personal Communication Magazine, vol. 6, no 2,pp 46-55, April 1999.
[3] D.Johnson , Y. Hu , D. Maltz," *The Dynamic Source Routing protocol (DSR) for Mobile Ad hoc network*", RFC 4728, 2007.
[4] S.Marti, T. Giuli, K.Lai, M.Baker, "*Mitigating routing Misbehaviour in Mobile Ad –hoc Networks*", In Proc of the Sixth International conference on Mobile Computing and networking(MOBICOM), Boston,2000.
[5] Abdelaziz Babakhouya, Yacine challal, "*A Simulation Analysis of Routing Misbehaviour in Mobile Ad-hoc Networks*", NGMAST-workshop on mobile security, Europe,2008.
[6] Fei Wang, Yijun Mo, Benxiong Huang, "*COSR: Cooperative On-Demand Secure Route Protocol in MANET*", IEEE conference ISCIT, pp. 890 – 893, China,2006.
[7] S. Zhong, J. Chen, and Y. Yang, "*Sprite: a simple, cheat-proof, creditbased system for mobile ad-hoc networks," IEEE INFOCOM 2003*, San Francisco, CA, USA, April 2003.
[8] NS2 network Simulator. http://www.isi.edu/nsnam/ns

AUTHORS PROFILE

Rekha Kaushik holds a Master of Technology(2008) from BarKatullah University , Bhopal, M.P. India and Pursuing Ph.D from MANIT Bhopal , India. She is a member of CSI and IETE (Institution of Electronics and Telecommunication Engineers). Her general research interests include wireless communication especially Ad-hoc network, network security.

Dr. Jyoti Singhai is Assistant professor in MANIT, Bhopal, India. She holds Ph.D degree from MANIT, India. Her general research interests include wireless communication , image processing , network security.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Survey on Fuzzy Clustering and Rule Mining

D.Vanisri
Department of Computer Technology
Kongu Engineering College
Perundurai, Tamilnadu, India
vanisri_raja@rediffmail.com

Dr.C.Loganathan
Principal, Maharaja Arts and Science College
Coimbatore, Tamilnadu, India
clogu@rediffmail.com

***Abstract*—The document clustering improves the retrieval effectiveness of the information retrieval System. The association rule discovers the interesting relations between variables in transaction databases. Transaction data in real-world applications use fuzzy and quantitative values, to design sophisticated data mining algorithms for optimization. If documents can be clustered together in a sensible order, then indexing and retrieval operations can be optimized. This study presents a review on fuzzy document clustering. This survey paper also aims at giving an overview to some of the previous researches done in fuzzy rule mining, evaluating the current status of the field, and envisioning possible future trends in this area**

Keywords- Fuzzy set, Fuzzy clustering, Fuzzy rule mining, Information Retrieval, Web analysis.

I. Introduction

Fuzzy sets used for optimization result by allowing partial memberships to the different sets. Fuzzy set theory provides the tools need to do the computations in order to be able to deal with different data structure. Data Mining is an analytic process designed to explore data in search of consistent patterns and systematic relationships between variables and then to validate the findings by applying the detected patterns to new subsets of data. The ultimate goal of data mining is extracting rules and clustering the similar objects.

The goal of this survey is to provide a comprehensive review of different fuzzy rule mining and clustering techniques in data mining. Clustering is a division of data into groups of similar objects. Each group, called cluster, consists of objects that are similar between themselves and dissimilar to objects of other groups. Representing data by fewer clusters necessarily loses certain fine details, but achieves simplification.

Association analysis is the discovery of what are commonly called association rules. It studies the frequency of items occurring together in transactional databases, and based on a threshold called support, identifies the frequent item sets. Another threshold, confidence, which is the conditional probability than an item appears in a transaction when another item appears, is used to pinpoint association rules. Association analysis is commonly used for market basket analysis. Clustering is the organization of data in classes. However, unlike classification, in clustering, class labels are unknown and it is up to the clustering algorithm to discover acceptable classes. Clustering is also called unsupervised classification, because the classification is not dictated by given class labels.

The remainder of this paper is organized as follows. Section II describes problem formation. Section III discusses some of the earlier proposed research work on fuzzy document clustering and fuzzy association rule mining. Section IV provides a fundamental idea on which the future research work focuses on. Section V concludes the paper with fewer discussions.

II. PROBLEM FORMULATION

Association Rule Mining (ARM) is the process of finding a rule of the form $X \cup Y$ from the given set of transactions. These transactions contain a set of items which is a subset of items in the set of unique items in the entire database. Association rule generated implies that if X, an item set specific to the domain is present then the probability of finding Y item set is given by confidence. The process of finding the association rules involves two steps namely frequent item set mining and association rule generation. Frequent item sets play an essential role in many data mining tasks that try to find interesting patterns from databases, such as association rules, correlations, sequences, episodes, classifiers, clusters and many more of which the mining of association rules is one of the most popular problems[1]. An association rule is an expression of the form $X \Rightarrow Y$, where X and

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Y are item sets, and $X \cap Y = \{\}$. Such a rule expresses the association that if a transaction contains all items in X, then that transaction also contains all items in Y. X is called the body or antecedent, and Y is called the head or consequent of the rule.To illustrate the concepts, for example from the supermarket domain.

The support of an association rule X => Y in D, is the support of X U Y in D, and similarly, the frequency of the rule is the frequency of X U Y. An association rule is called frequent if its support (frequency) exceeds a given minimal support (frequency) threshold σ. The confidence or accuracy of an association rule X => Y in D is the conditional probability of having Y contained in a transaction, given that X is contained in that transaction:

$$\text{confidence } (X \Rightarrow Y, D) = P(Y/X)$$
$$P(Y/X) = \text{support}(X \cup Y, D) / \text{support}(X, D)$$

The rule is called confident if P(Y|X) exceeds a given minimal confidence threshold γ, with $0 < \gamma < 1$. Based on classical association rule mining, a new approach has been developed expanding it by using fuzzy sets.

The clustering problem is expressed as follows: The set of N documents $D = \{D_1, D_2, ... D_N\}$ is to be clustered. Each $D_i \varepsilon U^{Nd}$ is an attribute vector consisting of N_d real measurements describing the object. The documents are to be grouped into non-overlapping clusters $C = \{C_1, C_2, ... C_N\}$ (C is known as a clustering), where, K is the number of clusters, $C_1 \cup C_2 \cup ... \cup C_K$, $C_i \neq \varphi$ and $C_1 \cap C_2 = \varphi$ for $i \neq j$.

Assuming f: $DxD \rightarrow U^+$ is a measure of similarity between document feature vectors. Clustering is the task of finding a partition $\{C_1, C_2, ..., C_K\}$ of D such that $\forall i, j \in \{1, ... K\}$, $j \neq i$, $\forall x \in C_i$: $f(x, O_i) \geq f(x, O_j)$ where, O_i is one cluster representative of cluster C_i.

The goal of clustering is stated as follows:
Given:

- A set of documents $D = \{D_1, D_2, ... D_N\}$
- A desired number of clusters K
- An objective function or fitness function that

evaluates the quality of a clustering, the system has to compute an assignment g: $D \rightarrow (1, 2, ..., K)$ and maximizes the objective function.

III. Related work

One of the key operations in fuzzy logic and approximate reasoning is the fuzzy implication, which is usually performed by a binary operator, called an implication function or, simply, an implication. M. Mas, et.al.,[2] tries to compile the main basic theoretical properties of the four most usual kinds of implications: S-, R-,QL-, and D-implications. This is done for the properly fuzzy environment (implications defined on [0,1]) as well as for the discrete case, which is increasingly studied because it allows to avoid numerical interpretations of the linguistic variables used in fuzzy techniques.

C.Y. Suen et al., [3] Handwriting recognition is a complex and important problem. Recognition of handwriting is important for automatic document processing functions such **as** mail sorting and check reading. Recognition of isolated handwritten digits is no longer a significant research problem. Paul D. Gader and James M. Keller [4] introduced fuzzy set theory to handwriting recognition and suggested a new application to handwritten word recognition.

Now-a-days, fraud prevention and detection is a very big category in research issues. Hence need some specific solutions and methodologies for preventing fraud. Mirjana[5] based on science database, fraud prevention has been conducted due to problem domains, fraud detection and prevention are diversified which is indicated by research articles survey. In this work, following applications areas were detected and described: telecommunications, insurance, auditing, medical care, credit card transactions, e-business, bid pricing and identity verification.

Fuzzy clustering is a widely applied method for obtaining fuzzy models from data. It has been applied successfully in various fields including finance and marketing. Fuzzy set theory was initially applied to clustering in [6]. The book by Bezdek [7] is a good source for material on fuzzy clustering. The most popular fuzzy clustering algorithm is the fuzzy c-means (FCM) algorithm. The design of membership functions is the most important problem in fuzzy clustering. Different choices include those based on similarity decomposition and centroids of clusters.

Eduardo Raul Hruschka et al., [8] gives survey on evolutionary algorithms for clustering. They proposed hard partition algorithms, though overlapping (soft/fuzzy) approaches and discussed key issues on the design of evolutionary algorithms for data partitioning problems, such as usually adopted representations, evolutionary operators, and fitness functions. In particular, mutation and crossover operators commonly described in the literature are conceptually analyzed, giving especial emphasis to those genetic operators specifically designed for clustering problems.

Chin-Teng Lin and Ya-Ching Lu,[9] Introduced a system, that has fuzzy supervised learning capability. With fuzzy supervised learning, it has been used for a

fuzzy expert system, fuzzy system modeling or rule base concentration. It has been also used for an adaptive fuzzy controller, when learning with numerical values.

Raghu Krishnapuram et al.,[10] presented new relational fuzzy clustering algorithms based on the idea of medoids. The worst case complexity of the algorithms was, which happens while updating the medoids in each iteration. This complexity compares very favorably with other fuzzy algorithms for relational clustering. These approach were useful in Web mining applications such as categorization of Web documents, snippets, and user sessions.

Chun-Hao Chen et al.,[11] put forward new view called cluster-based fuzzy-genetic mining algorithm for extracting both fuzzy association rules and membership functions from quantitative transactions. It can dynamically adjust membership functions by genetic algorithms and uses them to fuzzify quantitative transactions. It can also speed up the evaluation process and keep nearly the same quality of solutions by clustering chromosomes. Each chromosome represents a set of membership functions used in fuzzy mining. This algorithm first divides the chromosomes in a population into k clusters by using the k-means clustering approach. All the chromosomes in a cluster then use the number of large 1-itemsets derived from the representative chromosome in the cluster and their own suitability of membership functions to calculate the fitness values. The evaluation cost can thus be reduced due to the time-saving in finding 1-itemsets.

Hongwel Chen et.al [12], presented a general fuzzy trust problem domain for P2P-based system, and compare Fuzzy Comprehensive Evaluation method, Fuzzy Rank-ordering method, and Fuzzy Inference method through a concrete paradigm. In this paradigm, they had applied algorithm to Fuzzy Comprehensive Evaluation Method for P2P-based trust system, and Blin algorithm to that of Fuzzy Rank-ordering Method, and Mamdani algorithm to that of Fuzzy Inference Method. Results demonstrate that different fuzzy trust method for P2P-based system may deduce different fuzzy results.

Zhongze Fan and Minchao Huang, [13] specially makes extension of the conception of the fuzzy rule that the reasoning result may be any of all classes with different degrees though the premise is similar, thus the contradictions among the fuzzy rules can be completely resolved though there are overlaps among the hyper spheres. This idea can be applied for the fault diagnosis fields but also can be used for automata, signal treatment and image treatment etc.

FUZZY clustering techniques have been applied effectively in image processing, pattern recognition and fuzzy modeling. The best known approach to fuzzy clustering is the method of fuzzy -means (FCM), proposed by Bezdek [14] and Dunn [15], and generalized by other authors. A good survey of relevant works on the subject can be found in [16]. In FCM, membership functions are defined based on a distance function, and membership degrees express proximities of entities to cluster centers. By choosing a suitable distance function different cluster shapes can be identified [17]–[22]. Another approach to fuzzy clustering due to Krishnapuram and Keller [23] is the possibilistic-means (PCM) algorithm which eliminates one of the constraints imposed on the search for partitions leading to possibilistic (absolute) fuzzy membership values instead of FCM probabilistic (relative) fuzzy memberships. Usana Susana Nascimento et.al.,[24] introduced FCPM frame work called fuzzy clustering with proportional membership model, it says how data are generated from a cluster structure to be identified. This implies direct interpretability of the fuzzy membership values, which should be considered a motivation for introducing data-driven model-based methods. Hamid Mohamadlou et al., [25] spotted about an algorithm based on fuzzy clustering for mining fuzzy association rules using a combination of crisp and quantitative data. L. Bobrowski and J. Bezdek, [26], the reduction in the amount of clustering data allows a partition of the data to be produced faster.

Yücel Saygin and Özgür Ulusoy[27] forward to put some methods for automated construction of fuzzy event sets which are sets of events where each event has a degree of membership to a set. Fuzzy event sets are constructed by analyzing event histories. They have proposed a sliding window algorithm for mining event histories and proposed an automated rule modularization method that does not rely on semantic knowledge. Rafael Alcala et al., [28] based on the 2-tuples linguistic representation model, they have presented a new fuzzy data-mining algorithm for extracting both association rules and membership functions by means of an evolutionary learning of the membership functions, using a basic method for mining fuzzy association rules. Mila Kwiatkowska et al.,[29] reuse and integration of data from heterogeneous data sources requires explicit representation of the predictors, their measures, and their interpretations. They have described a new framework based on semantic and fuzzy logic for knowledge representation and secondary data analysis.

Yeong-Chyi Lee et al.,[30] gave an idea about multiple-level taxonomy and multiple minimum supports to find fuzzy association rules in a given quantitative transaction data set. Using different criteria to judge the importance of different items, managing taxonomic relationships among items, and dealing quantitative data sets are three issues that usually occur in real mining applications. This fuzzy mining algorithm can generate

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

large itemsets level by level and then derive fuzzy association rules from quantitative transaction data.

Yo-Ping Huang and Li-Jen Kao[31] introduced a model to find inter-transaction fuzzy association rules that can predict the variations of events. They proposed algorithm first mapped a quantitative attribute into several fuzzy attributes. A normalization process was taken to prevent the total contribution of fuzzy attributes from being larger than one. In order to mine inter-transaction fuzzy association rules, both the dimensional attribute and sliding window concepts were introduced in this approach.

Heng-Ming Huang projected a new fuzzy data-mining algorithm for extracting interesting knowledge from object-oriented quantitative transactions. The numbers of fuzzy intra-object association rules are usually smaller than those of fuzzy inter-object association rules because the attribute number is less than the item number in real applications. Finding inter-object association rules thus spends more time than finding intra-object association rules[32]. Tzung-Pei Hong[33] constructed several GA-based fuzzy data-mining methods for automatically extracting membership functions for the rules. All the genetic-fuzzy mining methods first use evolutional computation to find membership functions suitable for mining problems and then use the final best set of membership functions to mine fuzzy association rules.

IV. Future enhancement

In future we plan to investigate various ways of constructing the new fuzzy algorithms and apply different clustering methods. We will also apply a stability-based criterion for determining the optimal number of clusters. The topic of association rules has been studied over a decade. Most of the foundation researches have been done. A lot of attention was focus on the performance and scalability of the algorithms, but not enough attention was given to the interestingness of the rules generated. Although rule mining can help reveal patterns and relationships, it does not tell the user the value or significance of these patterns. These types of determinations must be made by the user. To enhance the performance of the system we can develop some intelligent system using fuzzy technique.

V. Conclusion

The volume of text data in the web is increasing exponentially, it makes difficult for searching purpose, several search engines in the web makes it possible to retrieve web documents by usual text database. However, users may not judge easily whether the documents have useful information, especially in the case that given keywords have wide concept, in order to .retrieve efficiently web documents, so here came the technique called "clustering". In this article, we discussed about introduction of field of fuzzy data mining. Therefore, we motivated this field of research, and gave more formal definition of the terms used and presented a brief overview of currently available fuzzy clustering and rule mining methods, their properties and their application to specific problems. Even though, it is impossible to describe all algorithms and applications in detail, but our ideas will be interesting to every reader to provoke for their further studies. We already know that "necessity is the mother of invention", while reading this paper, most of them can have lot of questions in them. This will strive path to have a new invention in the field of fuzzy data mining.

REFERENCES

[1]. Gosta Grahne and Jianfei Zhu (2006) "Efficiently Using Prefix-trees in Mining Frequent Itemsets" Concordia University Montreal, Canada, 2006.

[2]. M. Mas, M. Monserrat, J. Torrens, and E. Trillas, "A Survey on Fuzzy Implication Functions," IEEE Tranasaction On Fuzzy Systems, Vol. 15, 2007.

[3]. C. Y. Suen, C. Nadal, R. Legault, T. Mai, and L. Lam,"Computer Recognition of Unconstrained Handwritten Numerals," Proc.

[4]. Paul D. Gader and James M. Keller, "Applications of Fuzzy Set Theory to Handwriting Recognition, " Electrical and Computer Engineering Department , University of Missouri-Columbia.

[5]. Mirjana Pejic-Bach, "Profiling intelligent systems applications in fraud detection and prevention: survey of research articles," International Conference on Intelligent Systems, Modelling and Simulation, 2010, PP.80-85

[6]. RUSPINI, E. H., "A new approach to clustering", Inf Control 15, 1969, pp.22-32.

[7]. BEZDEK, J . C., "Pattern Recognition with Fuzzy Objective Function Algorithms", Plenum Press , New York, NY, 1981.

[8]. Eduardo Raul Hruschka, Ricardo J. G. B. Campello, Alex A. Freitas, and Andr´e C. Ponce Leon F. de Carvalho, "A Survey of Evolutionary Algorithms for Clustering," IEEE Transactions on Systems , Man and Cybernetics," Vol. 39, No. 2, March 2009 PP-133-155.

[9]. Chin-Teng Lin, and Ya-Ching Lu, " A Neural Fuzzy System with Fuzzy Supervised Learning," IEEE Transactions on Systems , Man and Cybernetic," Vol. 26, No. 5, October 1996.

[10]. Raghu Krishnapuram, Anupam Joshi, Olfa Nasraoui, and Liyu Yi, "Low-Complexity Fuzzy Relational Clustering Algorithms for Web Mining," IEEE Transactions On Fuzzy Systems, Vol. 9, No. 4, Auguest 2001, pp-595 -607.

[11]. Chun-Hao Chen, Tzung-Pei Hong, and Vincent S. Tseng," A Cluster-Based Fuzzy-Genetic Mining Approach for Association Rules and Membership Functions," IEEE International Conference on Fuzzy Systems July 2006, pp.16-

[12]. Hongwei Chen, ShengSheng Yu, Jianga Shang, Chunzhi Wang, Zhiwei Ye, "Comparison with Several Fuzzy Trust Methods for P2P-based System", IEEE International Conference on Information Technology and Computer Science, 2009 , PP-188-91.

[13]. Zhongze Fan and Minchao Huang, "Fuzzy Rule Set Based Engine Fault Diagnosis,", 2009 IEEE.

[14]. BEZDEK,J.C., "Pattern Recognition with Fuzzy Objective Function Algorithms", Plenum Press , New York, NY, 1981

 http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[15]. J. Dunn, "A fuzzy relative of the isodata process and its use in detecting compact, well-separated clusters," J. Cybernet., vol. 3, no. 3, pp. 32–57, 1974.
[16]. J. Bezdek, J. Keller, R. Krishnapuram, and T. Pal, Fuzzy Models and Algorithms for Pattern Recognition and Image Processing. Norwell, MA: Kluwer, 1999.
[17]. G. Gustafson and W. Kessel, "Fuzzy clustering with a fuzzy covariance matrix," in Proc. IEEE Conf. Decision Control, 1979, pp. 761–766.
[18]. R. Davé, "Fuzzy shell-clustering and applications to circle detection of digital images," Int. J. Gen. Syst., vol. 16, no. 4, pp. 343–355, 1990.
[19]. L. Bobrowski and J. Bezdek, "c-Means with l and l norms," IEEE Trans. Syst., Man, Cybern., vol. 21, pp. 545–554, Mar. 1991.
[20]. J. Bezdek, R. Hathaway, and N. Pal, "Norm induced shell prototype (NISP) clustering," Neural, Parallel, Scient. Comput., vol. 3, pp.431–450, 1995.
[21]. F. Höppner, F. Klawonn, R. Kruse, and T. Runkler, Fuzzy Cluster Analysis. New York: Wiley, 1999.
[22]. F. Klawonn and A. Keller, "Fuzzy clustering based on modified distance measures," in Proc. 3rd Int. Symp. Advances Intelligent Data Analysis, vol. 1642, J. Kok, D. Hand, and M. Berthold, Eds., 1999, pp. 291–301
[23]. R. Krishnapuram and J. Keller, "A possibilistic approach to clustering," IEEE Transaction on. Fuzzy System," vol. 1, pp. 98–110, Apr. 1993.
[24]. Susana Nascimento, Boris Mirkin, and Fernando Moura-Pires, "Modeling Proportional Membership in Fuzzy Clustering ," IEEE Transactions On Fuzzy Systems, Vol. 11, No. 2, April 2003, PP-173-186.
[25]. Hamid Mohamadlou, Reza Ghodsi , Jafar Razmi, Abbas Keramati "A method for mining association rules in quantitative and fuzzy data," 2009 IEEE ,PP.453-458
[26]. BEZDEK, J . C., "Pattern Recognition with Fuzzy Objective Function Algorithms", Plenum Press , New York, NY, 1981.
[27]. Yücel Saygin and Özgür Ulusoy, "Automated Construction of Fuzzy Event Sets and Its Application to Active Databases," IEEE Transactions on Fuzzy Systems, Vol.9, No.3, June 2001, pp.450-460.
[28]. Rafael Alcala and Jesus Alcald-Fdez and M.J. Gacto and Francisco Herrera, "Genetic Learning of Membership Functions for Mining Fuzzy Association Rules," 2007 IEEE.
[29]. Mila Kwiatkowska, M. Stella Atkins, Najib T. Ayas, and C. Frank Ryan, "Knowledge-Based Data Analysis: First Step Toward the Creation of Clinical Prediction Rules Using a New Typicality Measure," IEEE Transactions on Information Technology in Biomedicine, Vol.11, No.6, November 2007, DOI: 10.1109/TITB.2006.889693
[30]. Yeong-Chyi Lee, Tzung-Pei Hong, and Tien-Chin Wang "Mining Fuzzy Multiple-level Association Rules under Multiple Minimum Supports," IEEE International Conference on Systems, Man, and Cybernetics October 8-11, 2006, Taipei, Taiwan.
[31]. Yo-Ping Huang and Li-Jen Kao, "A Novel Approach to Mining Inter-Transaction Fuzzy Association Rules from Stock Price Variation Data," IEEE International Conference on Fuzzy Systems, 2005, pp-791-796.
[32]. Cheng-Ming Huang, Tzung-Pei Hong, and Shi-Jinn Hormg, "Simultaneously Mining Fuzzy Inter- and Intra-Object Association Rules," IEEE International Conference on Systems, Man, and Cybernetics October 8-11, 2006, Taipei, Taiwan, PP-2778-2783.
[33] Tzung-Pei Hong , "On Genetic-Fuzzy Data Mining Techniques ," IEEE International Conference on Granular Computing,2007, DOI: 10.1109/GrC.2007.160, PP-1-3.

Ms.D.Vanisri has received the Master of Science in Mathematics in 2001 from Madurai Kamaraj University. Then she completed her Master of Philosophy in Mathematics in the year 2003. She has presented many papers in national and international conferences and also guided many UG projects. Now she is doing research in the field of Fuzzy datamining at Mother Terasa Women's University, Kodaikannal. Currently she is working as a Lecturer in the school of Computer Technology and Applications, Kongu Engineering College,Tamilnadu.

Dr.C.Loganathan qualified basically with B.Sc and M.Sc in Mathematics in 1978 and 1980 respectively from Madras University and subsequently with M.Phil and Ph.D in Mathematics from Bharathiar University has served in various capacities as faculty member and Head of the Department of Mathematics at Kongu Engineering College, Perundurai for more than a decade. He is at present working as Principal, Maharaja Arts and Science College, Coimbatore. His unquenchable thirst for academic achievements had culminated in the publication of series of research papers, numbering more than 12 in the leading-referred national and international journals. As a research guide, he has produced many M.Phil and Ph.D candidates. He is a reviewer of many referred international journals. His areas of interest encompass Applied Mathematics, Control Theory, Numerical Methods, Quantitative Techniques and Neural Networks and fuzzy datamining. He has co-authored the books on "Quantitative Methods in Management, Engineering Mathematics I and Engineering Mathematics II".

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

An Agent Based Approach for End-to-End QoS Guarantees in Multimedia IP networks

A.Veerabhadra Reddy
Senior Lecturer, Department of ECE
Government Polytechnic for Women, Hindupur
veerabhadrareddyphd@gmail.com

Dr. D. Sreenivasa Rao
Professor, Department of ECE
JNTU CE, Hyderabad
dsraoece@yahoo.co.uk

Abstract— **Quality of Service (QoS) guarantees are important, if the network capacity is insufficient, particularly for real-time streaming multimedia applications such as voice over IP. Differentiated Services or DiffServ are the services of the original internet that prioritizes flows according to their service class and provides much better bandwidth utilization. Predicting the end-to-end behavior and acquiring the method by which individual routers deal with the type of service field is difficult and fairly appropriate. Moreover it becomes more difficult if a pakcet crosses two or more DiffServ clouds, before reaching its destination. In this paper, we propose a QoS mapping framework to achieve scalability and end-to-end accuracy in QoS, using a Policy Agent (PA) in every DiffServ domain. This agent performs admission control decisions depending on a policy database. It configures the ingress and egress routers to perform traffic policing and conditioning jobs. Moreover, it constructs the shortest path between a source and destination satisfying the QoS constraints Bandwidth and Delay. By simulation results, we show that our proposed approach attains high throughput with reduced packet loss when compared with the normal DiffServ architecture.**

Keywords-Quality of Service (QoS); Policy Agent (PA); DiffServ domain; QoS Route Selection; Packet loss, Throughput.

I. INTRODUCTION

A. IP Networks

A computer network made of devices that support the Internet Protocol is an IP network [1]. In Internet Protocol Suite, IP is the primary protocol in Internet Layer which has the task of delivering the packets from source to destination mainly based on their address.

B. Quality of Service (QoS) in IP

When compared with the achieved service quality, the traffic engineering term quality of service (QoS) will refer to the resource reservation control mechanisms in both the fields of computer networking and other packet-switched telecommunication networks.The abiltiy of the QoS is to provide different priorties to different applications, users or data flows or guaranteeing a certain level of performance to a data flow. For example, it guarantees required bit rate, delay, jitter, packet dropping probability and/or bit error rate. Quality of Service (QoS) guarantees are important, if the network capacity is insufficient, particularly for real-time streaming multimedia applications such as voice over IP. This is because it often requies the fixed bit rate and they are delay sensitive and also in networks where the capacity is a limited resource (Eg. Cellular data commumnication). QoS mechanisms are not required in the absence of network congestion [1]. QoS is the most important implementation consideration within a converged network. It is a networking term that specifies a guaranteed network data performance level. Practically, QoS is a mechanism to ensure that audio and video data pass through the network with minimum delay. IP voice or videoconferencing calls will be unreliable, inconsistent, and often unsatisfactory, if network QoS is poor [2].

C. Two solutions for Quality of Service guarantees

(i) Differentiated services (DiffServ)

Differentiated Services or DiffServ are the services of the original internet which maintains stateless property. Differentiated Services is a computer networking architecture which specifies a scalable,simple, and coarse-grained mechanism for classifying, managing network traffic and providing QoS guarantees on modern IP networks [1]. The basic of this architecture is to provide network resources between the traffic aggregates. DiffServ prioritizes flows according to their service class and provides much better bandwidth utilization [3]

(ii) Integrated services

Services that require stateful architecture of the internet are known as Integrated Services or IntServ [1]. This architecture specifies the elements to guarantee QoS on the networks and it is the basis of the reservation of network resources between the individual flows [3]. The main idea of the service is the resource reservation and admission control. [4]. Deterministic bandwidth and end-to-end delays to the individual flows can be offered by the IntServ. Moreover, it depends upon the admission control by placing strict resource reservations which guarantees the worst case situation [3]. The following are the categories of services in this architecture:

- Guaranteed Services
- Controlled-load Service

Guaranteed Services: It is estimated as the strongest allowable service in the environment of the internet so far. It has the ability to provide per flow bandwidth and delay

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

guarantees and it can assure that the packets will arrive within a selected delivery time [1].

Controlled-load Service: It allows the services poorly. It supports the applications which are highly sensitive to congested networks such as real time applications and these applications must tolerate small amounts of loss and delay. If an application uses this service, the performance will not be affected even when the network load is increased. The traffic will be provided with service similar to normal traffic in a network under light condition [1].

D. Problems or Challenges of QoS

Many things can happen to packets as they travel from origin to destination, resulting in the following problems as seen from the point of view of the sender and receiver:

When the packets travel from the source to destination, it experiences the following problems as seen from the point of view fo the sender and the receiver.

Dropped Packets: The routers may fail to deliver (drop) some packets when they arrive, if the buffers of the dropped packets are already full. Dependng on the state of the network, some of the packets or none or all the packets might be dropped. Thus it is not possible to forecast the packets.

Delay: For a packet it may take a long time to reach its destination, because it gets held up in long queues, or takes a indirect route to avoid congestion. Excessive delay can render an application such as VoIP or online gaming unusable, in some cases.

Jitter: Packets may reach the destination with different delays from the source. A packet's delay varies with its position in the queues of the routers along the path between source and destination.This position can vary and thus it cannot be predicted. This variation in delay is known as jitter [1].

Out-of-order Delivery: When a group of packets are routed, then different packets may take different route. Each of the packets results in different delay because the order of the packets are changed from the source to the destination. Special additional protocols are required to rearrange the out-of-order packets.

Error: When packets are transmitted along a route, it may be misdirected or combined together or corrupted. The receiver have to detect this and the request the sender to resend packets [1].

E. Problems in Differentiated and Integrated Services

Predicting the end-to-end behavior and acquiring the method by which individual routers deal with the type of service field is difficult and fairly appropriate. Moreover it becomes more difficult if a pakcet crosses two or more DiffServ clouds, before reaching its destination.

Simple over-provisioning is an inefficient solution for the internet traffic which is highly bursty. If the network is dimensioned to carry all traffic with traffic management, it will cost an order of magnitude more than a network dimensioned to carry the same traffic. The traffic management is used to prevent the collapse during the peaks.

Measuring the peak load is not possible. Since the TCP protocol requests more bandwidth as the loss rate decreases, it is not possible to measure the links to avoid end-to-end loss altogether, when sending a large file. On the other hand, increasing the capacity of one link causes loss on a different link.

By dropping the packets which are expended in carrying these packets until now through the network, the resources will be wasted. The bandwidth consumption at the congestion point and in the network is caused by retransmitting this traffic in many cases. The packets must be discarded as close to the edge of the network as possible, while Diffserv is often implemented throughout a network to minimize this waste.

The problem with IntServ is that many states must be stored in each router. It is difficult to keep the path of all the reservations because it works on the small scale. Thus the architrecture is not much familiar [1].

In this paper, we propose a QoS mapping framework to achieve scalability and end-to-end accuracy in QoS, using a Policy Agent (PA) in every DiffServ domain. This agent performs admission control decisions depending on a policy database.

II. RELATED WORK

Kazi Khaled Al-Zahid et al [5], have presented a strategy for ETE QoS management in IP networks based on the use of programmable software agents. They have proposed a QoS-based routing architecture to serve multi-constrain ETE high priority applications. According to their proposal, the users can be electronically specify their QoS requirement from the host application based on their preference. Although, their proposed system has some performance limitations, but as a whole it is flexible, because the routing functionality is completely done by the agents which works as complements with the existing technology.

Sergio Gonzalez-Valenzuala et al [6] have investigated an improvement by developing algorithms for determining the optimal multipoint-to-point (mp2p) routes through the use of mobile software agents. They have presented an mp2p routing scheme using a mobile intelligent agent system, called WAVES. The agents work in a highly distributed and parallel manner, cooperating to determine optimal routes in an mp2p connection scenario. This work aims at closing the gap between the theoretical routing research based on mobile agents, and practical routing requirements for real world networks that are likely to be deployed during the forthcoming years.

Yao-Nan Lien et al [7] have stated briefly an approach for the problem of QoS budget allocation which is deliberated in optimization for increasing resource usage efficiency. The end-to-end QoS controller in QoS coordination layer has the capability of global resource planning. It suggests that an end-to-end QoS controller will plan all resource provisions according to the traffic demands, and all the resource

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

allocation policy will be in accord with the planning. Their framework with simulation study demonstrates that it can indeed substantially increase the total number of network paths under constraints of end-to-end QoS requirements.

Daniel Schlosser et al [8] have proposed a simple interface as an abstraction of a network service based on the service oriented architecture approach. The approach considers QoS as the network functionality the user is mainly interested in and includes charging. They have shown how QoS guarantees for several parts of one connection can be consolidated into a QoS description for the complete service. Moreover, they have discussed options to measure the QoS and presented measurements exposing the quality of an available active measurement tool, Cisco IP SLA.

Lynda Zitoune et al [9] have presented a reactive control policy which adapts the source bit rate to the reserved resources in order to ensure performance guarantees for multimedia applications. Their proposed method called flatness based trajectory tracking deals with drastic traffic flow rate changes and limits the traffic in order to respect the time constraints. They have showed the contribution of the reactive control and the dynamic regulation using purely control theoretic approaches which stabilize the network and avoid undesirable oscillations for the transmission of such critical flows. By their work they have presented a performance analysis for such rate control mechanism, and illustrate its feasibility through its implementation on MPLS-TE control plane of SSFnet/Glass simulator.

Rick Whitner et al [10] have examined the issue of matching active measurements to the network's QoS configuration when monitoring a QoS-enabled IP network. Initially, they have illustrated the issue using common active measurement techniques. Then, they have examined approaches to matching active measurements to the network's QoS configuration. Finally, they presented their experiences in prototyping one approach.

III. NETWORK MODEL

We assume that a communication network can be modeled using a graph $G = (V, E)$ where V is the set of nodes which could be routers, servers or switches and E represents the set of edges or links of the network. For any consecutive nodes a, b, the link l_{ab} can be expressed for different parameters as:

$$H_l = H_{ab}$$
$$C_l = C_{ab}$$
$$D_l = D_{ab}$$
$$B_l = B_{ab}$$

Where H_l, is the hops, C_l is the cost, D_l is the delay and B_l is the bandwidth of the link l, where the link $l \in E$ is directly connected by $a \in V$ and $b \in V$. These parameters may occur in either nodes or edges. And these have either positive or non-negative impact over the communication network. QoS for different parameter can be expressed using the following relation.

$$B_p = MIN\{B_l \mid l \in P\} \tag{1}$$

$$D_p = \sum_{l \in P} D_l + \sum_{n \in P} D_n \tag{2}$$

$$C_p = \sum_{l \in P} C_l + \sum_{n \in P} C_n \tag{3}$$

Where, P is the path from source s to the destination d.

B_p is the bandwidth of the path P

D_p is the delay of the path P

C_p is the cost of the path P

The problem is to find a path between s to d, such that it would satisfy all QoS constraints from source to destination. The above constraints can be categorized in two groups: link constraints and path constraints. Path constrains again consist of two classes: additive and multiplicative. Serving application that requires both of these constraints simultaneously is yet an unsolved problem.

IV. PROPOSED AGENT BASED APPROACH

A. Design Overview

In this work, we propose a QoS mapping framework for both user and administrative policy, qualitative and quantitative QoS constraints over the internet's DiffServ domain. We consider the Policy Agent (PA) that depends on the local state information to satisfy the end user and do not consider any central mechanisms such as bandwidth broker or adaptive bandwidth scheme. Thus in our approach, according to the service and the requirement of the end user, multi constraint MQoS is used for QoS mapping requested in different degrees by user applications.

By assigning each packet with an appropriate QoS level, the QoS control takes place. In order to manage traffic according to the traffic conditioning agreement specified in the service level agreement (SLA), the application layer is responsible for producing the MQoS and sending it to the ingress of a DiffServ domain. The PA dynamically configured the necessary interface based on the requested traffic's source and destination information. Thus the traffic which is marked as the high priority will get the opportunity while the BE traffic is considered as low priority. The process is repeated for each node along the destination, if the PA satisfies the requested MQoS. Otherwise, a negative notification is sent to the source that current network is unable to meet the requested QoS constraints.

B. QoS Moniroting by the PA

TABLE I. QOS RESOURCE MATRIX

N	Bandwidth	Delay	Cost
N_0	e_{00}	e_{01}	E_{02}
N_1	e_{10}	e_{11}	E_{12}
N_2	e_{20}	e_{22}	E_{23}

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

The QoS monitoring at each node involves checking whether there are sufficient resources for meeting the MQoS. This is performed by the QoS Mapping Engine (QME) of the PA at the routers or switches. The QME contains a 2-D resource matrix shown in Table I that maps different network resource parameters with its neighbor routers entity. In Table I, N denotes the current visiting node that meets all the requested constraints of host application and N_0, N_1, N_2 are the attached neighbors of N. e_{ij} (where i is the router entity and j is the constraint) in the resource matrix denotes the value of constraints with the attached interface. The main advantage of using PA in admission control is to find a path with the requested QoS constraints.

The objective of the PA at any node i , is to check the consistency of the following relations to optimize the requested MQoS.

$$MIN(B_{li-1}) \geq B_{cons},\ where\ l_{i-1} \in P \tag{1}$$

$$\sum_{s}^{i-1} D_l \leq D_{cons} \tag{2}$$

$$\sum_{s}^{i-1} C_l \leq C_{cons} \tag{3}$$

During the path selection, if PA_i accepts the request from PA_{i-1} , then PA_i prohibits the RE (router entity) to accept any new resource request from others until the current request expires. If PA_i fails to qualify any one of the inequalities shown above, a reject reply is send to PA_{i-1} so that PA_{i-1} can trigger the routing algorithm to find the next alternate path to the destination.

From the above description, it is easy to understand that PA can support as many as QoS constraints as the application requires. Addition of a new constraint is just to include it in the resource matrix at each node so the QME can take care off the incoming MQoS request for new constraint.

C. QoS Route Selection

The following assumptions are made in order to attain optimal performance:

1. The existing link state protocol such as Open Shortest Path Protocol is used to obtain the topology information.

2. A hop by hop parameter optimization is considered to reach the final destination rather than considering the whole path.

3. The PA and the corresponding routing entity are closely integrated in such a way that if there any changes in the routing entity, then PA is informed to make necessary changes in the QoS resource matrix.

If the routing decisions are made in each router with local information by the PA, then it is referred as hop by hop routing. In our algorithm, we maintain a topological order such as bandwidth, delay, cost, etc. Thus, the important parameter is checked first by the PA.

Algorithm

Our algorithm takes a sub-optimal path search approach for the selected QoS constraints. The required input parameter is the MQoS which includes the source (s), destination (d) and multiple QoS criterion. In this algorithm, { N } denotes the set of nodes that are involved in the path $P(s,d)$, while V is the total number of nodes in entire network.

1. MQoS is applied from s to the next hop router through the primary shortest path on the routing table.

2. If PA_i accept the request then
   The RV is updated to {N}, where RV is the route vector.
   PA stores the partial route and no new QoS provisioning is accepted for the resources.
   End if

3. If MQoS reaches the d, then
   positive feedback is sent to the source in the reverse unicast path.
   End if

4. If the QoS monitoring fails, then
   PA_i sends the negative feedback to the last router that accepts the request.
   End if

5. When an infeasible link is encountered, PA searches for alternate paths for that can support the requested QoS constraints.

6. PA then bypasses the MQoS through the alternate path.

7. The PA can trace back when it faces infeasible link that fails to satisfy the requested QoS constraints.

8. The algorithm runs repeatedly until it finds any path to destination if there is any. MQoS can be send with current accumulated value of the constraints from the stored value at any time when it is looking for alternate paths.

To find the route using MQoS properly, it requires a set of standards which should be implemented in the PA. Low priority traffic that travels in the same route may experience delay due to the high precedence of QoS traffic. If PA wants to satisfy the requested QoS constraints, it adjusts the router to handle high priority traffic. Another advantage of this approach is, PA can dynamically tune only those sub interfaces where high priority traffic actually flows and others IP interface can be remain untouched. When a session is closed PA can readjust the router's state for usual operation according to RE's need. Moreover PA can send advance warning message to other same priority or low priority streams to inform them to choose either different path or slow down their transmission to avoid congestion and loss of transmission quality.

V. SIMULATION RESULTS

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

A. Simulation Model and Parameters

In this section, we examine the performance of our agent based QoS mapping approach with an extensive simulation study based upon the ns-2 network simulator [11]. We compare our results with the normal DiffServ architecture. The topology used in our experiments is depicted in Figure 1. As we can see from the figure, we have five senders and five receivers connected by a ingress router E1 and egress router E2 through a core router.

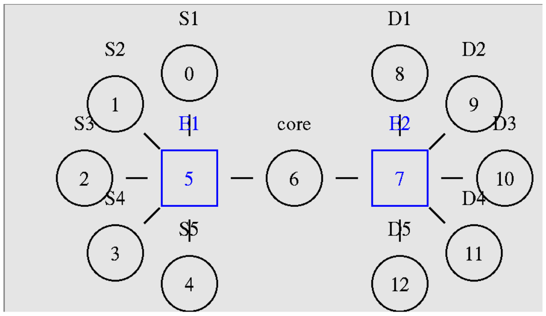


Figure 1. Simulation Topology

B. Performance Metrics

In our experiments, we vary the bottleneck bandwidth, traffic flow and traffic rate. We measure the following metrics:

- **Packet Loss**
- **Throughput in terms of packets**
- **Throughput in Mb/s**

The results are described in the next section.

C. Results

A. Effect of Varying Rate

In our first experiment, we vary the rate as 5Mb, 10Mb, 15Mb and 20Mb in order to calculate the packet loss, throughput (packets received) and throughput (Mbps). The results for the individual receivers are given.

1. Packet Loss

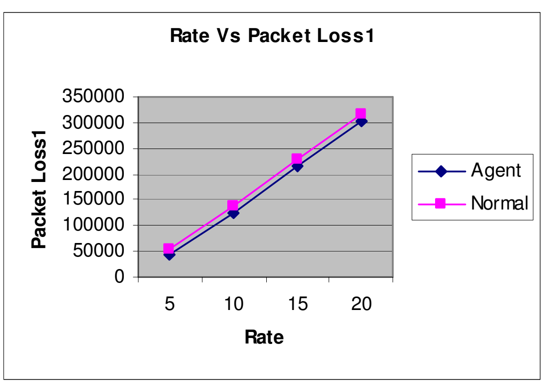


Figure 2. Rate Vs Packet Loss at Receiver 1

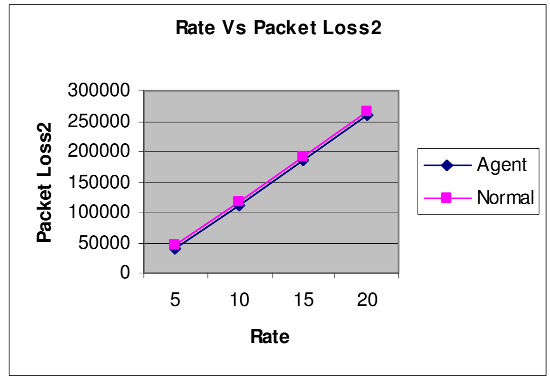


Figure 3. Rate Vs Packet Loss at Receiver 2

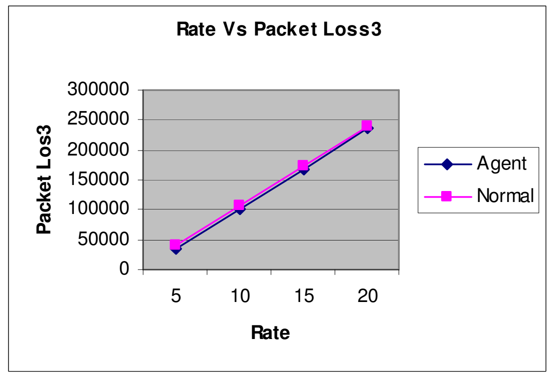


Figure 4. Rate Vs Packet Loss at Receiver 3

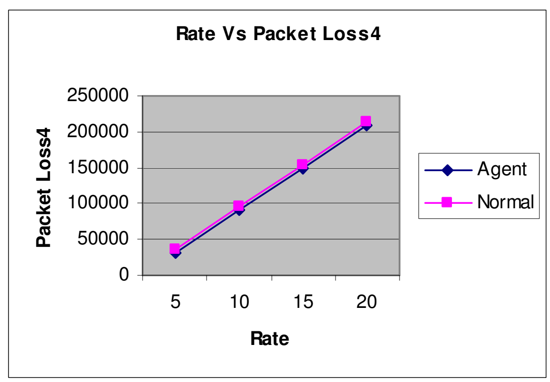


Figure 5. Rate Vs Packet Loss at Receiver 4

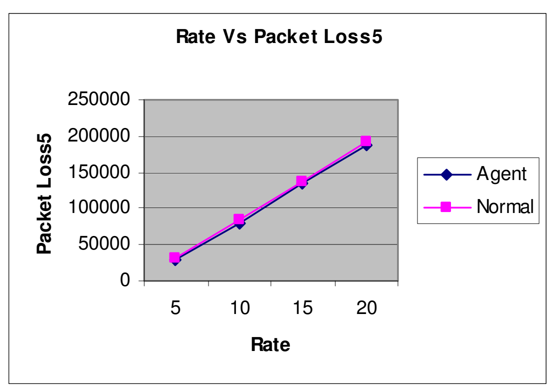


Figure 6. Rate Vs Packet Loss at Receiver 5

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Figure 2 to 6 shows the packet loss at the receivers 1 to 5 respectively. From the figure, we can see that the packet loss is high in the Normal scheme when compared with our Agent based scheme when varying the rates.

2. Throughput (Packets)

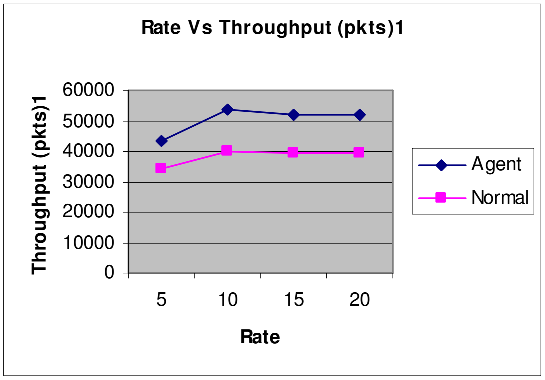


Figure 7. Rate Vs Throughput at Receiver 1

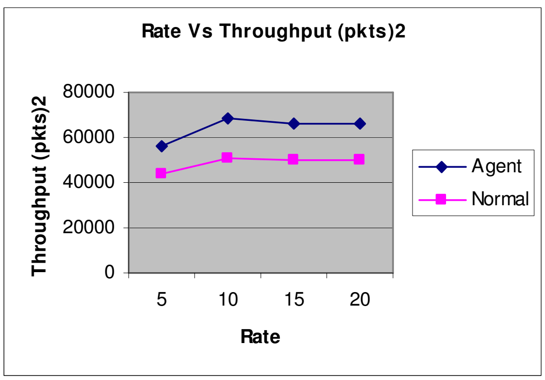


Figure 8. Rate Vs Throughput at Receiver 2

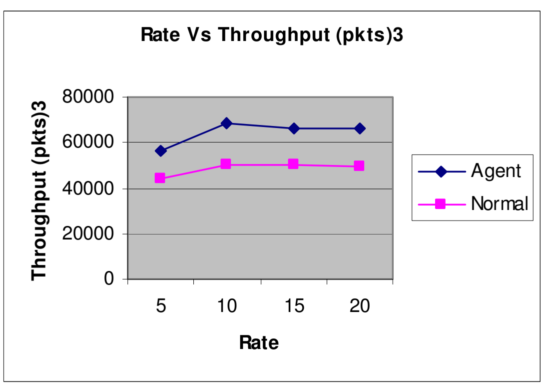


Figure 9. Rate Vs Throughput at Receiver 3

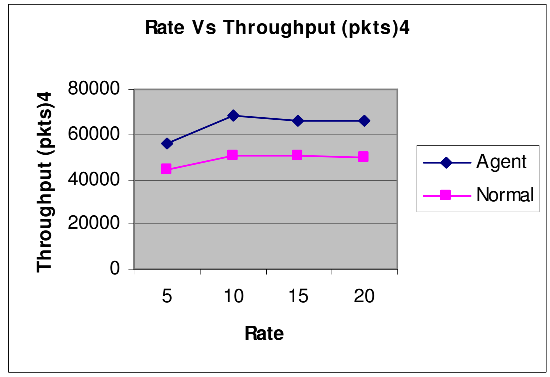


Figure 10. Rate Vs Throughput at Receiver 4

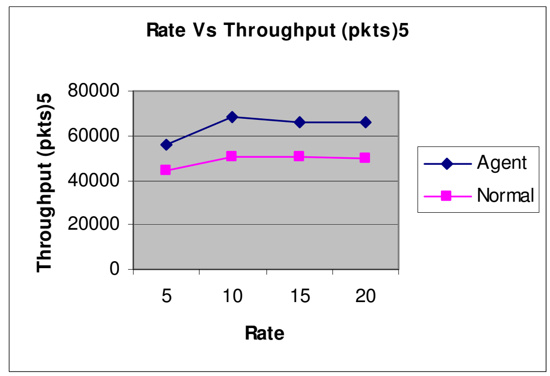


Figure 11. Rate Vs Throughput at Receiver 5

Figure 7 to 11 gives the Throughput in packets for the receivers 1 to 5 by varying the rates. It shows that the Throughput is more in the case of Agent based scheme when compared with Normal scheme

3. Throughput (Mbps)

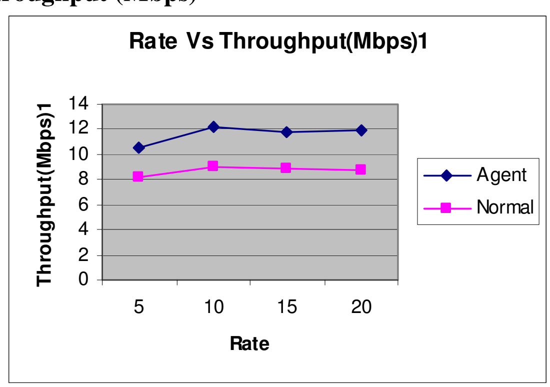


Figure 12. Rate Vs Throughput at Receiver 1

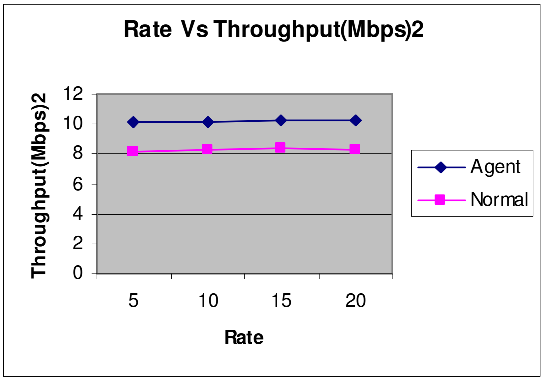


Figure 13. Rate Vs Throughput at Receiver 2

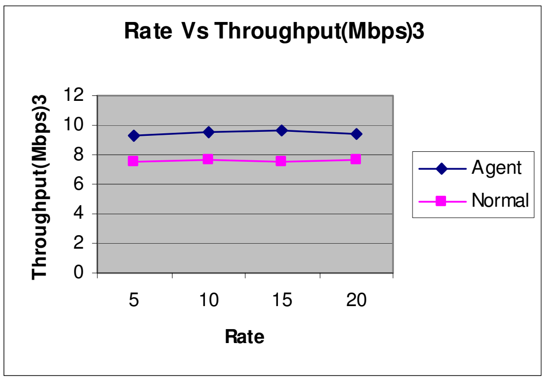


Figure 14. Rate Vs Throughput at Receiver 3

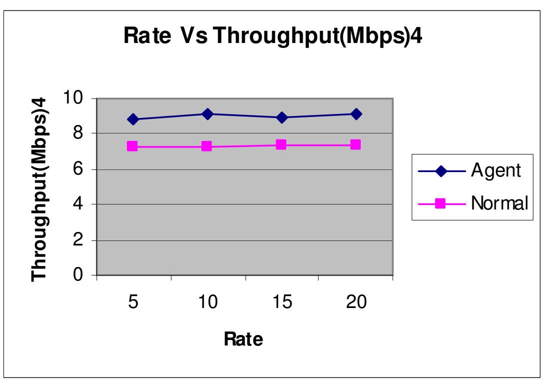


Figure 15. Rate Vs Throughput at Receiver 4

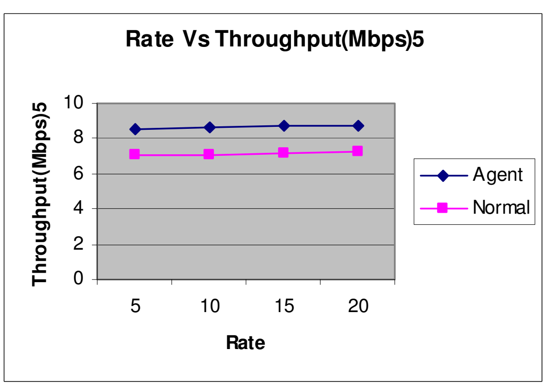


Figure 16. Rate Vs Throughput at Receiver 5

Figure 12 to 16 gives the Throughput in Mbps for the receivers 1 to 5 for varying the rates. It shows that the Throughput is more in the case of Agent based scheme when compared with Normal scheme.

B. Effect of Varying Simulation Time

In our second experiment, we vary the time as 2, 4,6,..10 seconds in order to calculate the packet loss, throughput (packets received) and throughput (Mbps). The results for the individual receivers are given.

1. Packet Loss

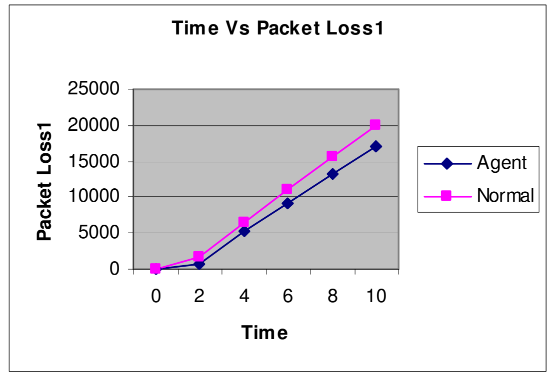


Figure 17. Time Vs Packet Loss at Receiver 1

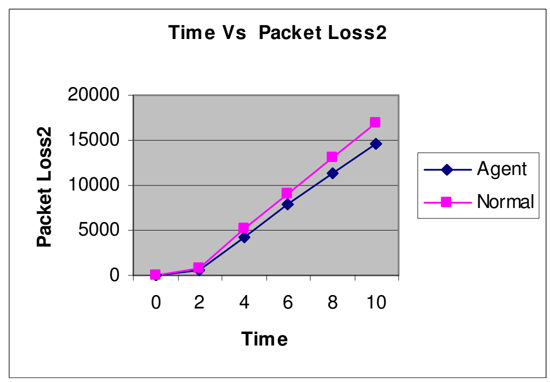


Figure 18. Time Vs Packet Loss at Receiver 2

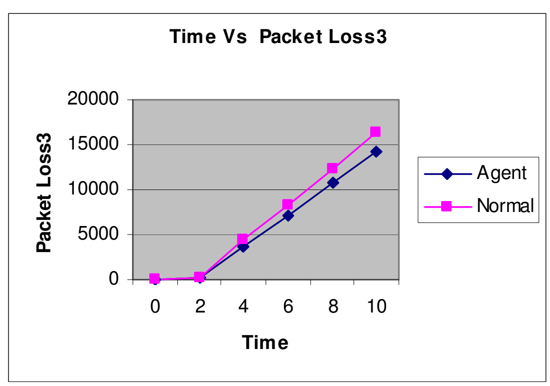


Figure 19. Time Vs Packet Loss at Receiver 3

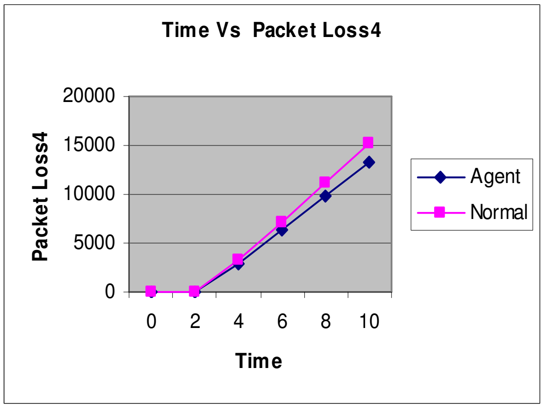


Figure 20. Time Vs Packet Loss at Receiver 4

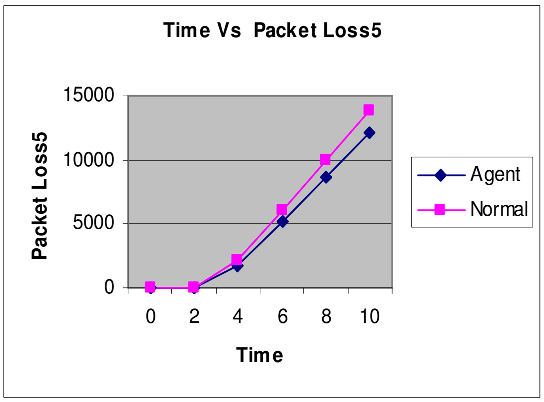


Figure 21. Time Vs Packet Loss at Receiver 5

Figure 17 to 21 show the packet loss for the receivers 1 to 5. From the figures, we observe that the loss is high in the Normal scheme when compared with our Agent based scheme when varying the time.

2. Throughput in Packets

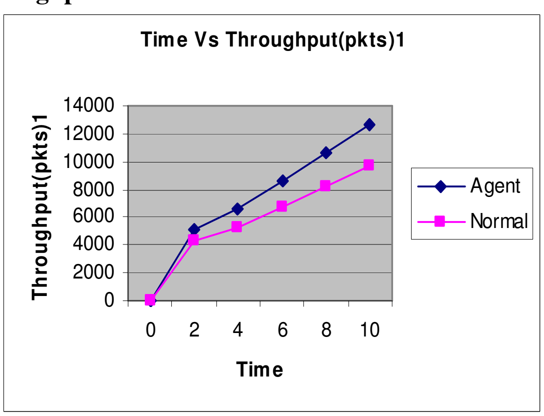


Figure 22. Time Vs Throughput at Receiver 1

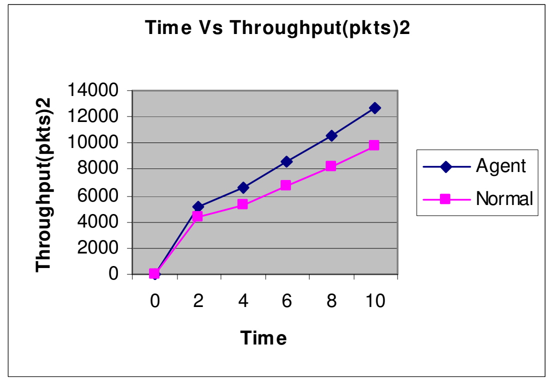


Figure 23. Time Vs Throughput at Receiver 2

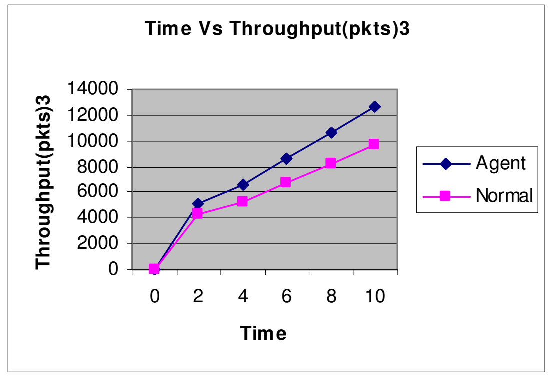


Figure 24. Time Vs Throughput at Receiver 3

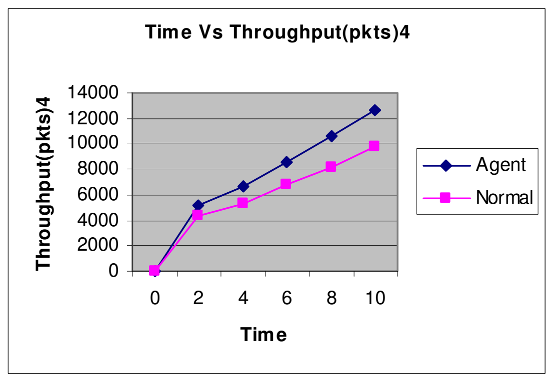


Figure 25. Time Vs Throughput at Receiver 4

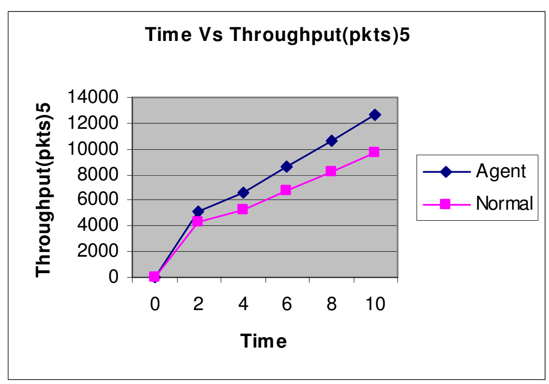


Figure 26. Time Vs Throughput at Receiver 5

Figure 22 to 26 give the Throughput in packets for the receivers 1 to 5 by varying the time. It shows that the Throughput is more in the case of Agent based scheme when compared with Normal scheme

3. Throughput (Mbps)

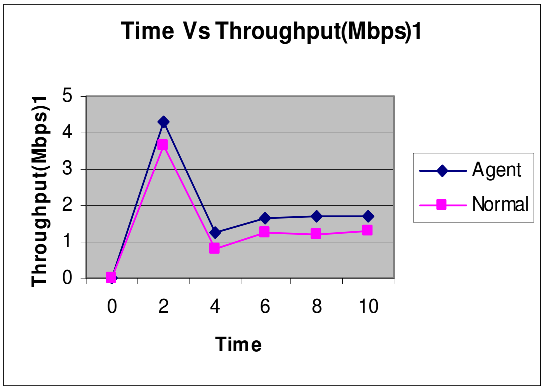


Figure 27. Time Vs Throughput Receiver 1

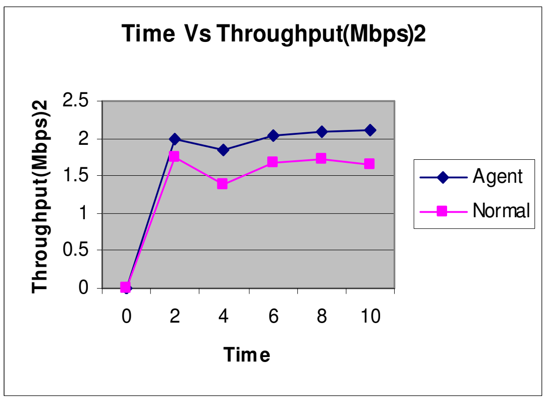


Figure 28. Time Vs Throughput Receiver 2

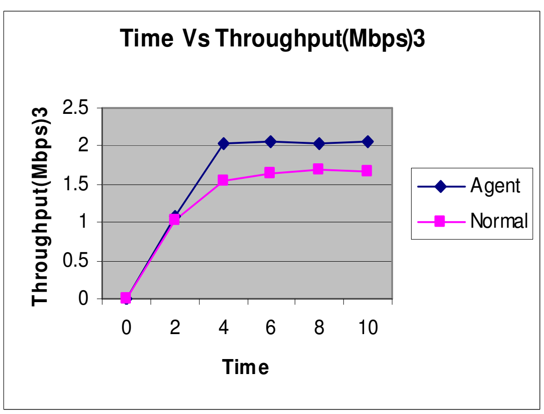


Figure 29. Time Vs Throughput Receiver 3

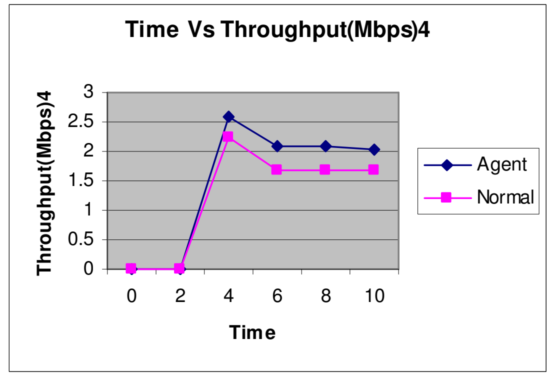


Figure 30. Time Vs Throughput Receiver 4

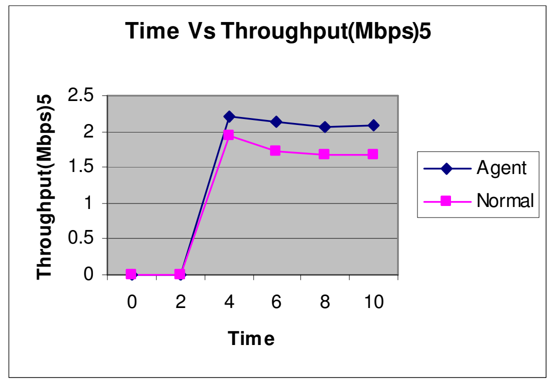


Figure 31. Time Vs Throughput Receiver 5

Figure 27 to 31 gives the Throughput in Mbps for the receivers 1 to 5 by varying the time. It shows that the Throughput is more in the case of Agent based scheme when compared with Normal scheme.

VI. CONCLUSION

In this paper, we propose a QoS mapping framework to achieve scalability and end-to-end accuracy in QoS, using a Policy Agent (PA) in every DiffServ domain. This agent performs admission control decisions depending on a policy database. It configures the ingress and egress routers to perform traffic policing and conditioning jobs. The QoS monitoring at each node involves checking whether there are sufficient resources for meeting the Multiple QoS constraints (MQoS). This is performed by the QoS Mapping Engine (QME) of the PA at the routers or switches. Moreover, it constructs the shortest path between a source and destination satisfying the QoS constraints Bandwidth and Delay. During the path selection, if PA at node i accepts the request from its previous node, then PA prohibits the router entity to accept any new resource request from others until the current request expires. If PA fails to qualify any one of the inequalities, a reject reply is send to the PA at previous node so that it can trigger the routing algorithm to find the next alternate path to the destination. By simulation results, we have shown that our proposed approach attains high throughput with reduced packet loss when compared with the normal DiffServ architecture.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

REFERENCES

[1] http://en.wikipedia.org/wiki

[2] E. Brent Kelly, "Quality of Service in Internet Protocol (IP) Networks", Infocomm – 2002.

[3] S. Terrasa, S. Saez, J. Vila and E. Hernandez, "Comparing the utilization bounds of IntServ and DiffServ", supported by the "HET-NETs – 2004.

[4] R.Braden, D.Clark and S.Shenker, "Integrated Services in the Internet Architecture: an Overview", RFC Editor, July 1994.

[5] Kazi Khaled Al-Zahid and Mitsuji Matsumoto, "Software Agent (SA) to guarantee QoS for multi constrain applications in all-IP networks", Second International Conference on Mobile Computing and Ubiquitous Networking, April 2005.

[6] Sergio Gonzalez-Valenzuela, Victor C. M. Leung and Son T. Vuong, "Multipoint-to-Point Routing With QoS Guarantees Using Mobile Agents", Mobile Agents for Telecommunication Applications, SpringerLink, January 2001, DOI: 10.1007/3-540-4651-6.

[7] Yao-Nan Lien, Hsing Luh and Chien-Tung Chen, "End-to-end QoS with Budget-Based Management", Proc. of the 2003 First International Working Conference on Performance Modeling and Evaluation of Heterogeneous Networks, July 2003.

[8] Daniel Schlosser and Tobias Hobfeld, "Service Oriented Network Framework Enabling Global QoS and Network Virtualization", 20th ITC Specialist Seminar, 18.-20. May 2009.

[9] Lynda Zitoune, Amel Hamdi, Hugues Mounier and Veronique Veque, "Dynamic Resource Management Approach In QoS-Aware IP Networks Using Flatness Based Trajectory Tracking Control", IEEE, IET International Symposium On Communication Systems, Networks And Digital Signal Processing, 2009.

[10] Rick Whitner, Graham Pollock and Casey Cook, "On Active Measurements in QoS-Enabled IP Networks", In PAM'02, Fort Collins CO, Mar. 2002.

[11] Network Simulator: www.isi.edu/nsnam/ns

A.Veerabhadra Reddy completed his B.Tech in Electronics and Communication Engineering from Bapatla college of Engineering In 1988.He worked as a production Engineer in Unitron ltd. Faridabad for one year and from June 1989 to June 1990 worked as Asst. professor in ECE at KITS, Ramtek. Then he has been serving to department of Technical Education A.P, Hyderabad from 1990. He completed his M.Tech (ECE) from JNTU, Kakinada in 2005. Now he is holding the post of Senior lecturer in ECE at Govt. polytechnic for women, Hindupur and additional charge to Govt. Polytechnic, Dharmavaram as an Offier on Special Duty. He was the visiting faculty to RGM Engineering College, Alfa College of Engineering, and Sri Ramakrishna post graduate college, Nandyal, A.P and taught various subjects in Computer Science and Electronics. He worked for 5 years as Assistant project officer in Community polytechnic scheme of MHRD, Govt. of India attached to polytechnics. He has been persuing his Ph.D under the guidance of Dr. D.Sreenivasa Rao, Professor, JNTU.

Dr.D.Srinivasa Rao has 20 years of teaching experience. He worked at CBIT as Lecturer in ECE Department for 6 years during 1988 – 1994. He worked at ECE Department of JNTU, Aanantapur in various capacities for 11 years during 1994-2005. Presently he is working as Professor in ECE Department of JNTU CE, Hyderabad. His research interest are in the area of communications and computer network s Presently 12 research students are working under his guidance. He has 22 publications in various National, International Conferences and Journals. He has attended more than 10 Short Term Courses, Summer Schools, and Workshops, conducted by various organizations. He has organized workshops and refresher courses. He has chaired sessions at various national conferences. He is advisory committee member for GNIT, Hyderabad. He is also governing body member for Syed Hashim College of Science & Technology, member of JNTU forum for Science & Society and Coordinator for campus networking at JNTU CE,

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

High Performance Reconfigurable Balanced Shared Memory Architecture For Embedded DSP

J.L.Mazher Iqbal
Assistant Professor,
ECE Department,
Rajalakshmi Engineering College,
Chennai-602 105, India
mazheriq@gmail.com

Dr.S.Varadarajan
Associate Professor,
ECE Department,
Sri Venkateswara College of Engineering,
Sri Venkateswara University,
Tirupati-517 502, India
varadasouri@gmail.com

***Abstract*—Reconfigurable computing greatly accelerates a wide variety of applications hence it has become a subject of a great deal of research. It has the ability to perform computations in hardware to increase performance, while keeping much of the flexibility of a software solution. In addition reconfigurable computers contain functional resources that may be easily modified after field deployment in response to changing operational parameters and datasets. Till date the core processing element of most reconfigurable computers has been the field programmable gate array (FPGA) [3]. This paper presents reconfigurable FPGA-based hardware accelerator for embedded DSP. Reconfigurable FPGAs have significant logic, memory and multiplier resources. These can be used in a parallel manner to implement very high performance DSP processing. The advantages of DSP design using FPGAs are high number of Instructions/Clock, high number of Multipliers, high Bandwidth Flexible I/O and Memory Connectivity. The proposed processor is a reconfigurable processing element architecture that consists of processing elements (PEs), memories and interconnection network and control elements. Processing element based on bit serial arithmetic (multiplication and addition) was also given. In this paper, it is established that specific universal balanced architecture implemented in FPGA is a universal solution, suited to wide range of DSP algorithms. At first the principle of modified shared-memory based processor are shown and then specific universal balanced architecture is proposed. An example of processor for TVDFT Transformation on the given accelerator is also given. By the proposed architecture, we could reduce cost, area and hence power in the best-known designs in the Xilinx FPGA technology.**

Key Word; Reconfigurable architectures; FPGA; Pipeline; Processing Element; Hardware Accelerator

I. INTRODUCTION

Now that design rules have stopped shrinking for ASICs, ASSPs and the like, they seem likely to be replaced by FPGAs. With their design rules coming into the 40nm generation, FPGAs will soon be level with ASICs and ASSPs in terms of circuit size and performance. The circuit configuration of FPGAs can be freely revised by equipment manufacturers on the spot, which means that they have no need to pay development costs, including mask sets. Even better, FPGAs do not require any circuit fabrication after design, which means faster equipment development. Nowadays, consumer appliances have become more advanced than ever. They are required to be more functional and portable. Moreover, the span of the product's life has become shorter. There are two important issues to develop LSIs, period and cost. Development of new LSI demands investing hugely and taking a big risk. Programmable devices, such as CPU, DSP and FPGA, have become a key to resolve these issues and hardware reconfigurability has been paid attention because of its high performance [3]. FPGA has high flexibility and suitable to implement control circuits, but FPGA suffers from the low area efficiency to implement data dominated circuits. When implementing industrial application systems, the area of FPGA implementation is far larger than that of ASIC implementation because of the high reconfigurability. Reconfigurable architecture has the capability to configure connections between programmable logic elements, registers and memory in order to construct a highly parallel implementation of the processing kernel at run time. This features makes them attractive, since a specific high speed circuit for given instance of an application can be generated at compile or even run time. Since the appearance of the first reconfigurable computing systems, DSP applications have served as important test cases in reconfigurable architecture and software development. In the area of special purpose architecture for digital signal processing, systolic arrays are recognized as a standard for high performance**.** Systolic designs represent an attractive architectural paradigm for efficient hardware implementation of computation-intensive DSP applications, being supported by the features like simplicity, regularity and modularity of structure. In addition, they also possess significant potential to yield high-throughput rate by exploiting high-level of concurrency using pipelining or parallel processing or both [1]. Today's objective is to tailor system performance to given task at minimal cost in terms

of chip area and power consumption. Finding a universal solution suited to wide range of DSP algorithms is permanently actual task. To reach relevant real-time performance, it must be multiprocessor architecture. At the architectural level, the main interest is the overall organization of the system compound using processing elements (PEs), memories, communication channels and control elements. One of the possible approaches is so called shared memory architecture. Our architecture can obtain a high area efficiency and high performance for implementing industrial applications.

II. PRINCIPLE OF SHARED MEMORY BASED PROCESSOR

In this section, we review shared-memory approach for DSP application [13]. The architecture of shared memory is shown in Figure 1. The idea is very simple. In order to simultaneously provide the PEs with input data, we need to partition the shared-memory into blocks. Processing elements (PEs) usually perform simple memory less mapping of the input values to a single output value. Using a rotating access scheme, each processor gets access to the memories once per N (N - number of PE's) cycles. During this time processor either writes or reads data from memory. All processors have the same duration time slot to access to the memories and access conflict is completely avoided. The disadvantage of using shared-memory architecture is the memory bandwidth bottleneck. In order to avoid bandwidth bottleneck and simultaneously provide the processors with several (K) input data, the shared-memory is partitioned into K memories (figure 3).

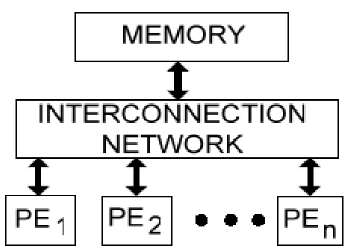


Fig. 1. Shared-memory architecture

In this paper, a special instance of that architecture is presented. The main target is to find balance between complexity of interconnection network, type of computation model of PEs (serial vs. parallel), number of PEs and memory size. Chosen compromise should fulfill following factors: required performance, minimal power consumption and cost in terms of chip area. Another important requirement is to create flexible, easy reconfigurable architecture suited to wide range of DSP algorithms.

A. *Processing Elements (PEs)*

Usually Processing elements perform simple memory less mapping of the input values to a single output values. The PEs can be in parallel or serial fashion. In parallel form it requires a parallel data bus and careful design because of delays and carry propagation. Parallel form leads to arithmetic operation made in one clock cycle, but when compared to serial form, it consumes more chip area. Serial PEs receives their inputs bit serially, and their results are also produced bit-serially. Hence, only a single wire is required for each signal. Design process could be more simple and robustness. The cost in terms of chip area and power consumption is therefore low. However, to achieve required performance bit-serial communication leads to high clock frequencies.

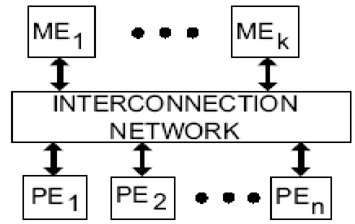


Fig. 2. Shared-memory architecture

B. *Memory elements*

Memory elements comparing to PEs are slow. It is desirable to make a trade-off between additional registers and RAM to achieve appropriate (in comparing to PE) read and write performance. By bit parallel PE high speed register will play a trivial (one word) cache memory role. By bit serial PE there must be a shift register. The data can be shifted in to and out from register with high speed. Then word can be written into the RAM. The RAM addressing requires only cyclic work. Reading data is bit-parallel and stored into shift register. Number of RAM words should be enough to store all variables accordingly to realize algorithm.

C. *Interconnection network (ICN)*

Interconnection network provides the communication channel needed to supply the PEs with proper data and parameters, and store results in the proper memories. The data movement should be kept simple, regular and uniform. Major design issues involve the topology of the communication network and its bandwidth.

III. BALANCED MODIFIED SHARED MEMORY ARCHITECTURE

Realizing single, basic arithmetic operation like addition or multiplication, it is obviously bit-parallel version of PEs that has several times higher performance then bit serial one. However, taking into account whole module with PEs, input and output registers, memory and interconnection network, advantage of parallel form is not so clear. Including power consumption and chip area, serial form could be more convenient. Generally smaller chip area and smaller clock leads to smaller power consumption. The requirements on the PE are that it completes its operation within the specified time limit. Self–explanatory chip area of single serial-PE is much smaller then parallel-PE, but to get the same performance needs faster clock. Parallel PE's lead to more connections lines, consistently more area and power. Parallel-PE looks to have several times bigger computational throughput (then serial by the same clock), however when considering 2-3 PE's in shared– memory

architecture it could be impossible to use high speed clock because of noise in signal propagation in long parallel buses. Otherwise control part of whole system in serial-PE version may be in micro program fashion, where implemented algorithm will be changed by the way of changing control memory contest. Using parallel-PE, control part must be significantly changed due to change type of computational task. This brief considerations show that shared-memory architecture with serial-PE's can be easier to implement and is better suited to wide range of DSP algorithms.

The proposed balanced shared-memory module based on the approach [13] is shown on figure 3. The Processing elements are capable of performing three computing functions: bit–serial full addition (inc. carry), bit–serial multiplication and negation, with two inputs and one output. Other arithmetic operation will be done as sequence of additions. Because of two serial inputs and one serial output, each PE is equipped with four shift registers (two inputs and one output for two independent memories). Those registers are used as single word cache memory and as serial/parallel parallel/serial translators on communication path to RAM memory. Hence interconnection network is very simple. This leads to small chip area and possibly of using high speed clock. Number of PEs should be 2 and consequently 1 RAM memory blocks (one RAM per multiplied output of PE). The multiplier output of PE1 is shipped into the RAM block using signals s1=1 and s2=0 to PE2 via shift register. PE2, accumulate the multiplier output and write back the result to the output buffer. Our shared-memory architecture offers good balance in terms of chip area, power consumption, computational throughput and flexibility. In fact of lack of required performance proposed module could be "multiplied" i.e. connected as shown in figure 5 and figure 6. Using parallel (figure 5) or cascade (figure 6) connection of modules it is easy to create processor suited to wide range of algorithms. Almost every required performance could be achieved as well. In the next part of this article the example realization of multiplier based on serial arithmetic and TVDFT transformation based on proposed architecture is presented. The heart of proposed architecture is PE (Figure 4). This element makes all bit-serial arithmetic calculation such as multiplication and addition. Moreover that, the processor element can made negation on "b" input when control signal "not_b" is high. Control signals s1 and s2 enable multiplication and addition respectively.

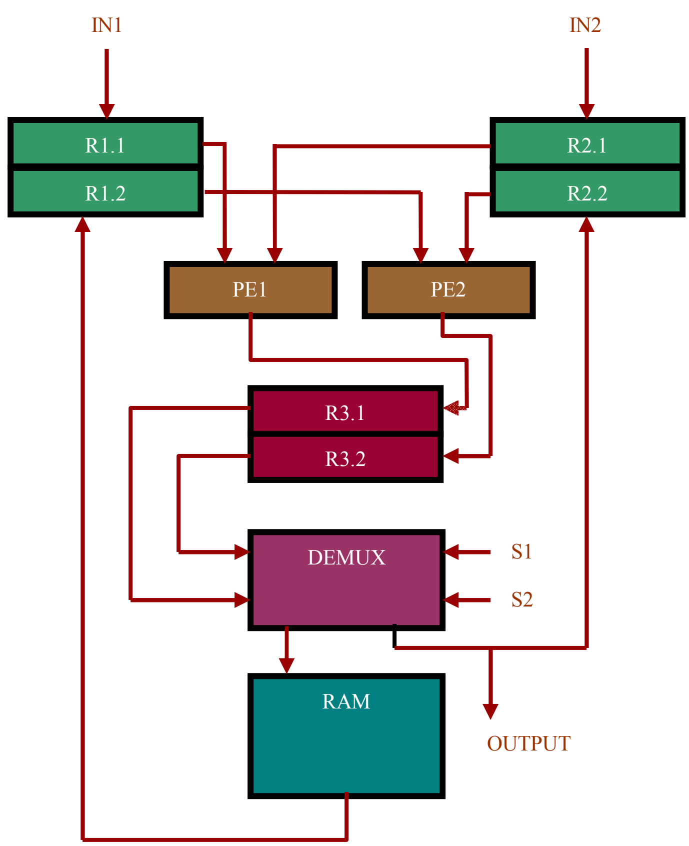


Fig. 3. Instance of universal specific balance architecture

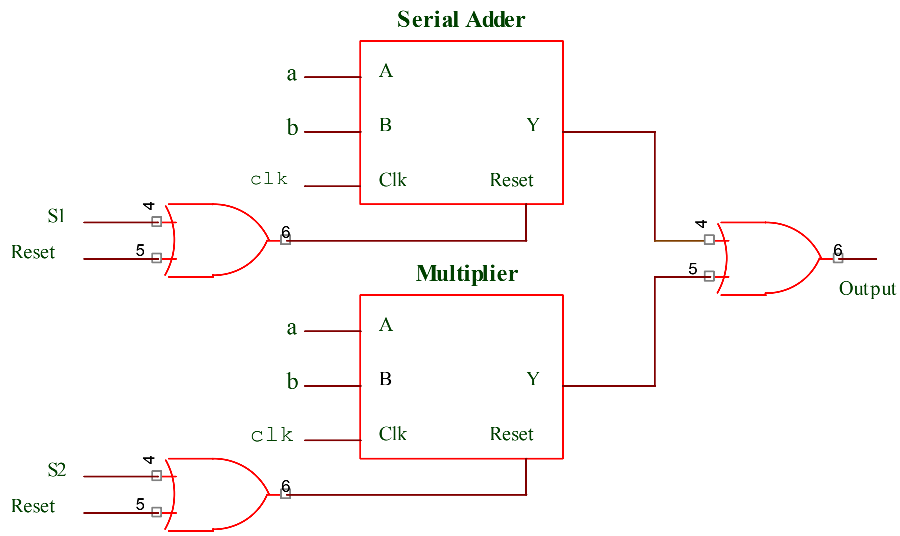


Fig. 4. PE architecture

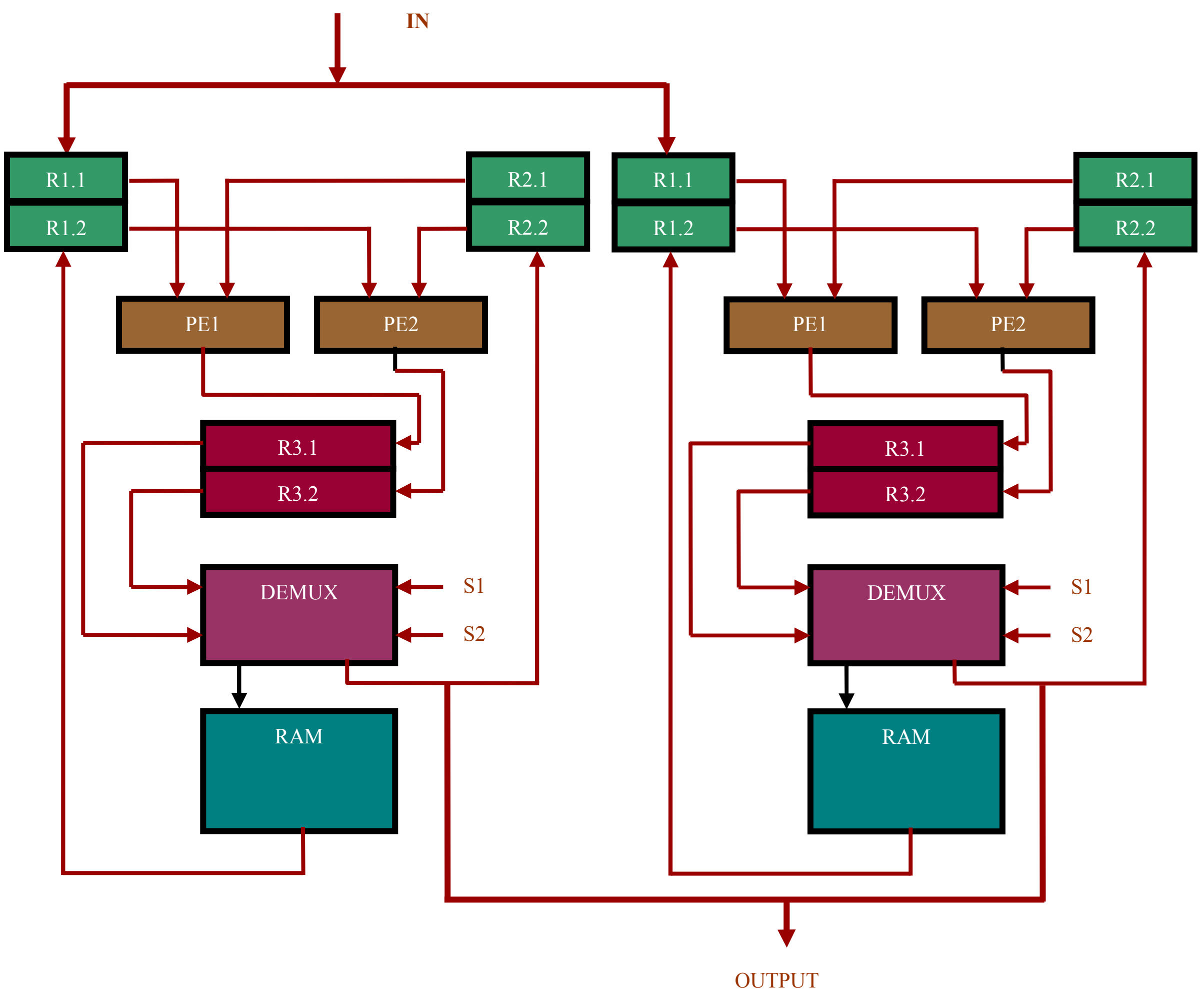


Fig. 5. Parallel form

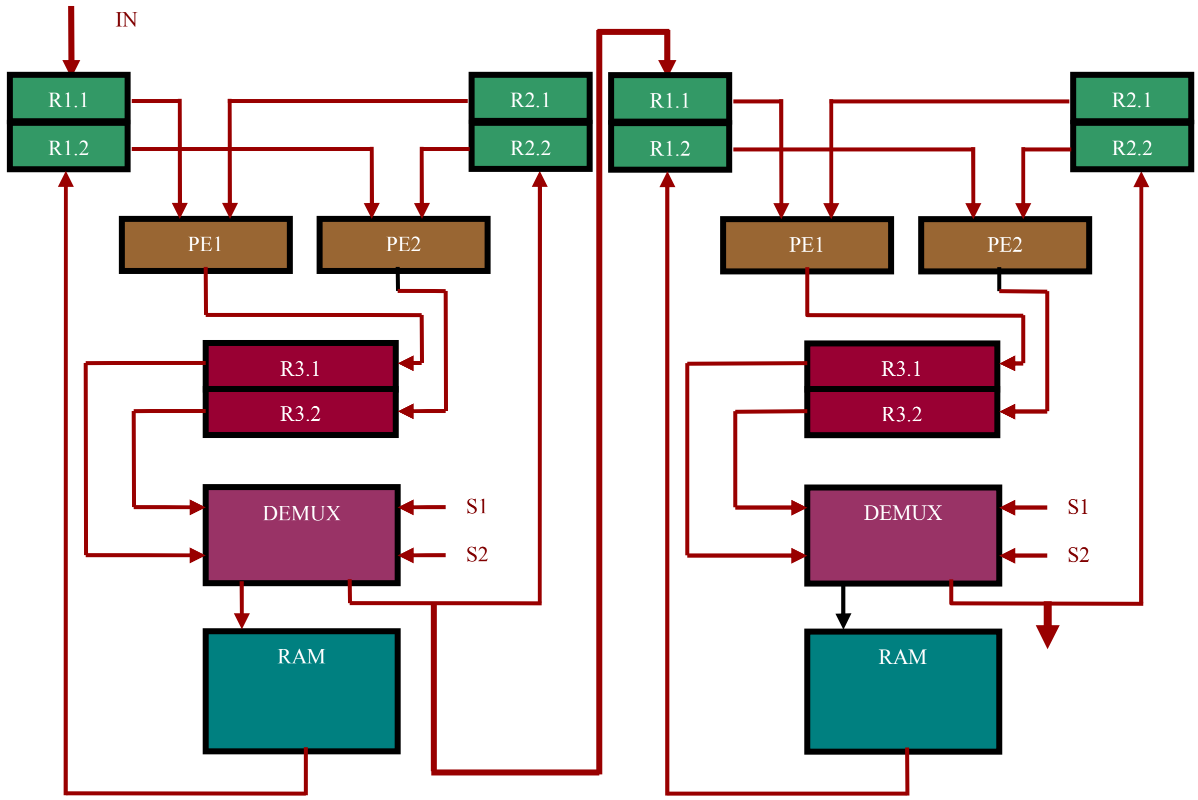


Fig. 6. Cascade form

A. *Multiplicator based on the serial arithmetic*

Most bit-serial multipliers are in practice based on the shift-and-add algorithm where several bit-products are added in each time slot. We present such bit-serial multiplicator. Bit-serial addition is based on equations are given bellow:

$$
\begin{aligned}
&Sum_i = XOR\{A_i, B_i, D_i\} = A_i \oplus B_i \oplus D_i \\
&C_i = Majority\{A_i, B_i, D_i\} \\
&\quad = A_i.B_i + A_i.D_i + B_i.D_i \\
&Di = C_{i-1} \text{ for } i = W_d - 1, W_d - 2, \ldots 1, 0 \qquad (1) \\
&D_{Wd-1} = 0 \\
&\text{Where}: W_d - \text{data word length}
\end{aligned}
$$

In bit–serial arithmetic the numbers are normally processed with least-significant bit first. In bit–serial carry–save adder carries of the adder are saved from one bit position to the next. At the beginning of computation the D flip-flop is reset. The n-bit serial multiplicator is shown on figure 7. At the first step register (shift register with serial input and parallel output) takes less significant bit of multiplier and first delay from element D1 (less significant bit of multiplicand). Unit delay D1 holds bits for one clock cycle and D2 unit, two cycles respectively.

The first bit (less significant) of output appears after 2n clock cycles. For example 8–bit serial multiplication takes 16 clock cycles.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

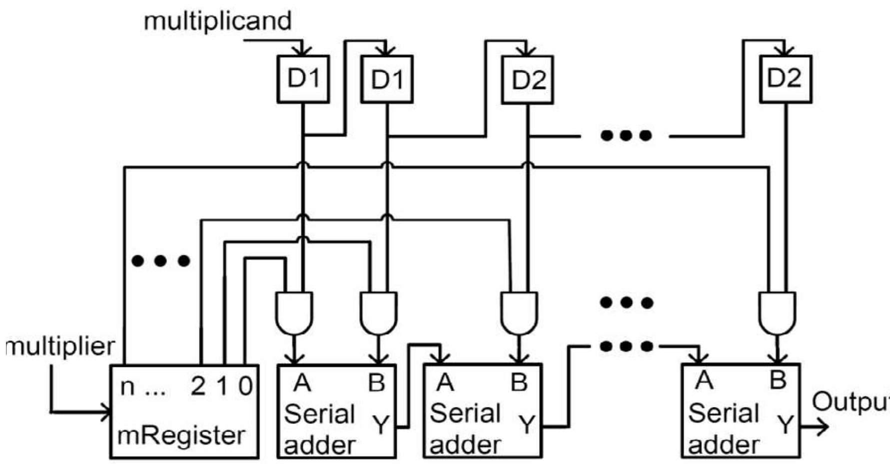


Fig. 7. Bit serial multiplicator

IV. TVDFT TRANSFOMATION ON THE GIVEN ARCHITECTURE

Systolic system consists of an array of processing elements (typically multiplier-accumulator chips) in a pipeline structure that is used for applications such as image and signal processing. Systolic approach can speed up a compute-bound computation in a relatively simple and inexpensive manner. A systolic array in particular achieves higher computation throughput without increasing memory bandwidth. A wide variety of signal processing functions can be hosted on the shared-memory processor, including complete subsystems that encompass multiple algorithms. The TVDFT transformation [8] will be used as the example. TVDFT is given by equation:

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j\varphi(n,k)} w(n), k = 0..k \qquad (2)$$

Where X (k) - spectral component corresponding to *k*-th harmonic, *N* – length of analysis frame, *x*(*n*) -input signal, *w*(*n*) – time window, K – number of orders in input signal.

$$\begin{cases} \varphi(0,k) = 0 \\ \varphi(n,k) = \sum_{i=1}^{n} \dfrac{2\pi(f_0(i) - f_0(i-1)),}{2Fs} \end{cases} \qquad (3)$$

Where f_0 (i) is fundamental frequency at time specified by i, F_s is sampling frequency.

In case of linear change of fundamental frequency formula (3) can be written as follows:

$$\varphi(n,k) = \frac{2\pi nk}{F_s}\left(f_0 + \frac{n\Delta f}{2N}\right) \qquad (4)$$

Where: f_0 -fundamental frequency at the beginning of analysis frame, Δf is fundamental frequency change within analysis frame. Hence, TVDFT formula (2) can be written as follows:

$$\begin{cases} \text{Re} \quad X(k) = \sum_{n=0}^{N-1} x_w(n) \cos \dfrac{2\pi nk}{F_s}\left(f_0 + \dfrac{n\Delta f}{2N}\right) \\ \text{Im} \quad X(k) = \sum_{n=0}^{N-1} x_w(n) \sin \dfrac{2\pi nk}{F_s}\left(f_0 + \dfrac{n\Delta f}{2N}\right) \end{cases}, \qquad (5)$$

where: xw(n)=x(n)w(n).

Formula (5) shows, that for practical realization of TVDFT, two sine wave generators with linear change of frequency can be used [10]. The balance for that algorithm needs two serial PEs, so there is only six shift registers for communication between RAM and PEs. Size of memories for given example is 3 of 8–bit words RAM [0...2]. Such architecture can be used as bit–serial computational unit (figure 3). According to proposed TVDFT computation in formula 5, at the first step we must calculate value of sine and cosine function then multiply that value by input signal. Given algorithm could be realized in one bit–serial computational unit in two main steps. Step one – generation of sine/cosine value [10].

As the main step of TVDFT algorithm there is necessary multiply generated value of sine/cosine by value of input signal. It could be done by detailed steps given bellow:

1*) Load data into input registers.*

- R1.1< -IN1, R2.1< - IN2;
- R1.2< -RAM [0], R2.2< -R3.2;

 http://sites.google.com/site/ijcsis/
ISSN 1947-5500

2) Calculations.

- PE1< -R1.1 * R2.1;
- PE2< -R1.2 + R2.2;

3) Writing the results into output registers.

- R3.1< -PE1;
- R3.2< -PE2;

4) Writing the result from output registers to memories

- RAM [0] < -R3.1;
- R2.2< -R3.2;

Where: IN1- input signal from sine generation out module, IN2- input signal X(n).

As it was shown, scheduling of TVDFT algorithm consists of two parts of work. First of it, computation value of sine/cosine function and the second one, multiplication of 8 bit input sample by 8 bit sine/cosine value, according formula 5. The same main steps should be repeated N times for each real and imaginary part. For improving computations performance it is possible to use parallel or cascade connection of computation units.

V. HARDWARE IMPLEMENTATION

FPGA implementation of specific universal balanced architecture was made on XILINX VIRTEX-II family (Selected Device: 2vp2fg256-6). Simulation is critical in verifying developed design behavior. Functional and timing simulation of the specific universal balance architecture design was done with Mentor Graphics Modelsim using developed test bench and appropriate stimuli to validate the design. Table I shows area utilization of given architecture and it is compared with utilization of area in previous architecture [1] [3] [13] [14]. Table II shows timing summary and Table III shows timing constraint. The simulation result of TVDFT Transform on specific universal balanced architecture is shown in figure 8. Layout design is created using Micro wind 3.1 to verify the impact of physical structure on the behavior of the developed design which is shown in figure 9. The final voltage, maximum Idd current of proposed architecture is 0-1v 0-2mA.

TABLE I. Device utilization summary

2vp2fg256-6	**Area Used**	**Utilization**
Number of Slices	40 out of 1408	2%
Number of Slice Flip Flops	48 out of 2816	1%
Number of 4 input LUTs	72 out of 2816	2%
Number of bonded IOBs	28 out of 140	20%
Number of MULT18X18s	2 out of 12	16%
Number of GCLKs	1 out of 16	6%

TABLE II. Timing summary

Maximum Frequency	271.444MHz
Minimum Period	4.160ns
Minimum input arrival time before clock:	3.316ns
Maximum output required time after clock	3.670ns

A. Comparison of Architectures

Details of the performance of the specific universal balanced architecture of Section III in terms of the basic design metrics are tabulated alongside with those of other comparable existing architectures in Table IV and figure 10. It is clear that the proposed implementation significantly outperforms the existing implementations in terms of three important key metrics, namely the area occupied, maximum usable frequency, and gate count.

TABLE III. Timing constraint: Default period analysis for Clock 'clk'

Cell:in->out	**Fan out**	**Gate Delay**	**Net Delay**	**Logical Name (Net Name)**
MULT18X18S:C>P4	2	0.705	0.561	PE1/multiplier1/Mmult__old_pdt_int_11_inst_mult_0
LUT2_D:I1->O	6	0.313	0.524	PE1/output4<4>1 (PE_1<4>)
MULT18X18S:A4		2.057		PE2/multiplier1/Mmult__old_pdt_int_11_inst_mult_0
Total		4.160ns		(3.075ns logic, 1.085ns route)(73.9% logic, 26.1% route)

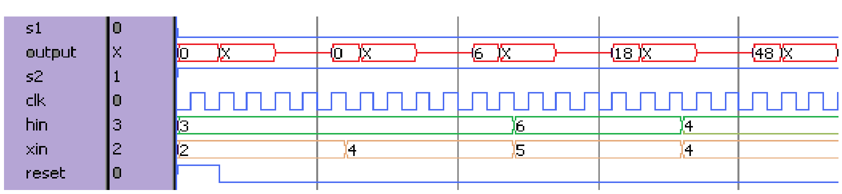


Fig. 8. Simulation of instant universal specific balance architecture

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Table IV. Comparison of Performance of the Proposed Implementation and the Existing Reconfigurable Implementation

2vp2fg256-6	**Area Used**			
	Proposed	**Systolic Fir Filter[14]**	**2-D Systolic Structure for FIR Filters[1]**	**Yoo et[13]**
Number of Slices	40	122	133	146
Number of Slice Flip Flops	48	668	48	48
Number of 4 input LUTs	72	-	--	--
Frequency[MHz]	271.44	84.5	74	70
Period[ns]	4.160	11.8	14.5	14.0

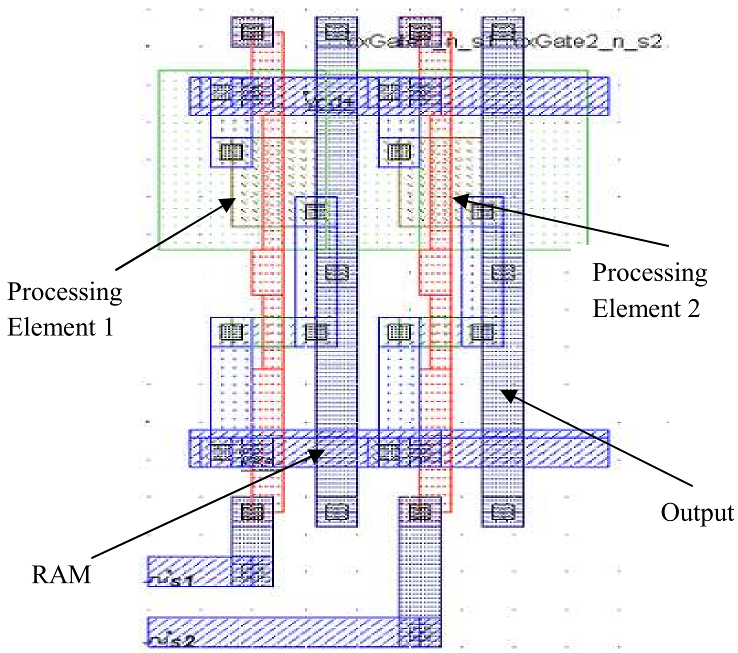


Fig. 9. Layout design of specific instant universal balance architecture

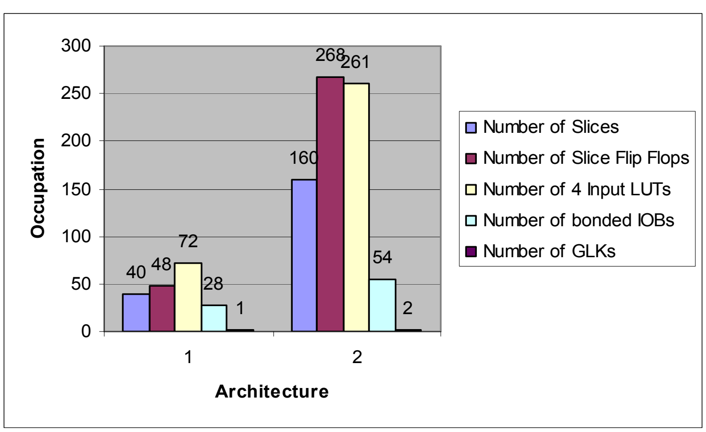


Fig. 10. Comparison of Proposed Architecture and Existing Shared Memory Architecture [3]

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

VI. CONCLUSION

Reconfigurable hardware is getting more and more complex with increased complexity and heterogeneity. To develop smaller and powerful reconfigurable processor we have proposed the universal computation module as balanced architecture, based on modified shared-memory approach. The balance was achieved between processing elements (PEs), count of memories and interconnection network. As an example of balanced architecture appliance for DSP algorithm the TVDFT transformation was implemented. Processing element based on bit serial arithmetic (multiplication and addition) was also given. As presented in this paper shared-memory balanced architecture implemented in FPGA is a universal solution, suited to wide range of DSP algorithms.

REFERENCES

[1] Pramod Kumar Meher, Abbes Amira, July 2008. "FPGA Realization of FIR Filters by Efficient and Flexible Systolization Using Distributed Arithmetic", IEEE Trans, Signal Processing, vol. 56, no. 7, pp. 3009-3017.

[2] Khaled Benkrid, Feb 2008. " High Performance Reconfigurable Computing: From Applications to Hardware", in IAENG International journal of computer science, 35:1, IJCS_35_1_04.

[3] G.Rubin, M. Omieljanowicz, A. Petrovsky, Jun 2007. "Reconfigurable FPGA-Based hardware accelerator for embedded DSP", in Proc MIXDES'07 Conf, pp.147-151.

[4] Makoto Okada, Tatsuo Hiramatsu, Hiroshi Nakajima, Makoto Ozone, Katsunori Hirase and Shinji Kimura, 2005. " A Reconfigurable Processor based on ALU array architecture with limitation on the interconnect", in Proc IEEE IPDPS'05, pp.1-6.

[5] H. Yoo and D. V. Anderson, "Hardware-efficient distributed arithmetic architecture for high-order digital filters," in Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP), Mar. 2005, vol. 5, pp. v/125–v/128.

[6] Multiprocessor systems-on-chips, 2005 edited by A.A.Jerraya, W.Wolf, Elsevier Inc.

[7] Jae-Jin Lee and Gi-Yong Song, "Implementation of a Bit-level Super-Systolic FIR Filter", 2004 IEEE Asia-Pacific Conference on Advanced System Integrated Circuits(AP-ASIC2004)I Aug. 4-5, 2004, pp.206-209.

[8] A. Petrovsky, P. Zubrycki, A. Sawicki, 2003. "Tonal and noise separation based on a pitch synchronous DFT analyzer as a speech coding method", The proc. of the ECCTD' 03, vol.III, Cracow, pp. 169-172.

[9] Katherine Compton, June 2002. "Reconfigurable Computing: A survey and software", in ACM Computing survey, Vol.34, N0.2, pp.171-210.

[10] M. Omieljanowicz, P. Zubrycki, A. Petrovsky, G.Rubin, 2002. "FPGA-based algorithms and hardware for generating digital sine waves", MIXDES 2002, Wroclaw, pp. 279-284.

[11] R.Tessier, W.Burleson, 2002. "Reconfigurable computing and digital signal processing: past, present, and future", in the book "Programmable digital signal processors: architecture, programming, and applications", edited by Yu Hen Hu, Marcel Dekker, Inc., pp.147-185

[12] R.Haritenstein, "Reconfigurable Computing: the Road map to a New Business Model-and its Impact on SOC Design," SBCCI 2001- 15th Symposium on Integrated Circuits and System Design, Brasilia, DF, Brazil, Sep.2001.

[13] L. Wanhammar, DSP integrated circuits, Academic Press, USA, 1999.

[14] R. Reeves, K. Sienski, C. Field, "Reconfigurable Hardware Accelerator for Embedded DSP", ICSPAT 97', SanDiego, 1997, 929-933.

[15] http://www.xilinx.com

AUTHORS PROFILE

J.L.Mazher Iqbal received B.E degree in Electronics and Communication Engineering from Madurai Kamaraj University in 1998 and M.E degree in Applied Electronics from Anna University in 2005. He is persuing Ph.D in Sri Venkateswara University College of Engineering, Sri Venkateswara University Tirupati. Currently, he is working as Assistant Professor in the Department of Electronics and Communication Engineering, Rajalakhmi Engineering College, Chennai, Tamil Nadu, India.

Dr. S. Varadarajan received B.Tech degree in Electronics and Communication Engineering from Sri Venkateswara University, Tirupati in 1987 and M.Tech degree from NIT, Warangal in Instrumentation in 1981, respectively. He obtained Ph.D. from Sri Venkateswara University, Tirupati in 1997. He is a fellow of IETE and member, IEEE. Currently, he is working as Associate Professor in the Department of Electronics and Communication Engineering, Sri Venkateswara University College of Engineering, Sri Venkateswara University, Tirupati, Andhra Pradesh, India.

A Novel approach of Data Hiding Using Pixel Mapping Method (PMM)

Souvik Bhattacharyya , Lalan Kumar and Gautam Sanyal

Abstract—Steganography is a process that involves hiding a message in an appropriate carrier like image or audio. The carrier can be sent to a receiver without any one except the authenticated receiver only knows existence of the information. Considerable amount of work has been carried out by different researchers on steganography. In this work the authors propose a novel Steganographic method for hiding information within the spatial domain of the gray scale image. The proposed approach works by selecting the embedding pixels using some mathematical function and then finds the 8 neighborhood of the each selected pixel and map each two bit of the secret message in each of the neighbor pixel according to the features of that pixel in a specified manner.This approach can be modified for mapping of four bits of the secret message by considering more no of features of the embedding pixel. Before embedding a checking has been done to find out whether the selected pixel or its neighbor lies at the boundary of the image or not. This solution is independent of the nature of the data to be hidden and produces a stego image with minimum degradation.

Keywords—Cover Image, Pixel Mapping Method (PMM), Stego Image.

I. INTRODUCTION

STEGANOGRAPHY is the art and science of hiding information by embedding messages within other, seemingly harmless messages. Steganography means "covered writing" in Greek. As the goal of steganography is to hide the presence of a message and to create a covert channel, it can be seen as the complement of cryptography, whose goal is to hide the content of a message. Another form of information hiding is digital watermarking, which is the process that embeds data called a watermark, tag or label into a multimedia object such that watermark can be detected or extracted later to make an assertion about the object. The object may be an image, audio, video or text only. A famous illustration of steganography is **Simmons' Prisoners' Problem** [16].An assumption can be made based on this model is that if both the sender and receiver share some common secret information then the corresponding steganography protocol is known as then the secret key steganography where as pure steganography means that there is none prior information shared by sender and receiver. If the public key of the receiver is known to the sender, the steganographic protocol is called public key steganography [2], [3] and [10].For a more thorough knowledge of steganography methodology the reader may see [14], [17].Some Steganographic model with high security features has been presented in [4], [5] and [6].Almost all digital file formats can be used for steganography, but the image and audio files are more suitable because of their high degree of redundancy [17]. Fig. 1 below shows the different categories of steganography techniques.

Fig. 1. Types of Steganography

A block diagram of a generic image steganographic system is given in Fig. 2.

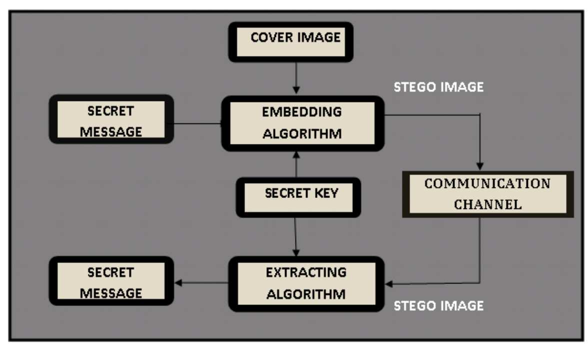


Fig. 2. Generic form of Image Steganography

A message is embedded in a digital image (cover image) through an embedding algorithm, with the help of a secret key. The resulting stego image is transmitted over a channel to the receiver where it is processed by the extraction algorithm using the same key. During transmission the stego image, it can be monitored by unauthenticated viewers who will only notice the transmission of an image without discovering the existence of the hidden message. In this work a specific image based steganographic method for gray level image has proposed. In this method instead of embedding the secret message into the cover image a mapping technique has been incorporated to generate the stego image. This method is capable of extracting the secret message without the presence of the cover image.

This paper has been organized as following sections: Section II describes some related works, Section III deals with proposed method. Algorithms are discussed in Section IV and Experimental results are shown in Section V. Section VI

S. Bhattacharyya is with the Department of Computer Science and Engineering, University Institute of Technology, The University of Burdwan, West Bengal, India e-mail: (souvik.bha@gmail.com).

L. Kumar is with the Central Institute of Mining and Fuel Research , Dhanbad, Jharkhand, India e-mail:(lalan.cimfr@gmail.com).

G. Sanyal is with with the Department of Computer Science and Engineering, National Institute of Technologyy West Bengal, India e-mail: (nitgsanyal@gmail.com).

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

contains the analysis of the results and Section VII draws the conclusion.

II. Related Works

A. Data Hiding by LSB

Various techniques about data hiding have been proposed in literatures. One of the common techniques is based on manipulating the least-significant-bit (LSB) [8], [9] and [13], [15]planes by directly replacing the LSBs of the cover-image with the message bits. LSB methods typically achieve high capacity but unfortunately LSB insertion is vulnerable to slight image manipulation such as cropping and compression.

B. Data Hiding by PVD

The pixel-value differencing (PVD) method proposed by Wu and Tsai [18] can successfully provide both high embedding capacity and outstanding imperceptibility for the stego-image. The pixel-value differencing (PVD) method segments the cover image into non overlapping blocks containing two connecting pixels and modifies the pixel difference in each block (pair) for data embedding. A larger difference in the original pixel values allows a greater modification. In the extraction phase, the original range table is necessary. It is used to partition the stego-image by the same method as used to the cover image. Based on PVD method, various approaches have also been proposed. Among them Chang et al. [12]. proposes a new method using tri-way pixel-value differencing which is better than original PVD method with respect to the embedding capacity and PSNR.

C. Data Hiding by GLM

In 2004, Potdar et al.[11] proposes GLM (Gray level modification) technique which is used to map data by modifying the gray level of the image pixels. Gray level modification Steganography is a technique to map data (not embed or hide it) by modifying the gray level values of the image pixels. GLM technique uses the concept of odd and even numbers to map data within an image. It is a one-to-one mapping between the binary data and the selected pixels in an image. From a given image a set of pixels are selected based on a mathematical function. The gray level values of those pixels are examined and compared with the bit stream that is to be mapped in the image.

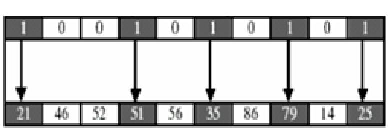


Fig. 3. Data Embedding Process in GLM

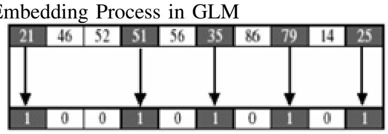


Fig. 4. Data Extraction Process in GLM

D. Data Hiding by the method proposed by Ahmad T et al.

In this work [1] a novel Steganographic method for hiding information within the spatial domain of the grayscale image has been proposed. The proposed approach works by dividing the cover into blocks of equal sizes and then embeds the message in the edge of the block depending on the number of ones in left four bits of the pixel.

III. Proposed Method

In this section the authors propose a new method for information hiding within the spatial domain of any gray scale image.This method can be considered as the improved version of [7].The input messages can be in any digital form, and are often treated as a bit stream. Embedding pixels are selected based on some mathematical function which depends on the pixel intensity value of the seed pixel and its 8 neighbors are selected in counter clockwise direction. Before embedding a checking has been done to find out whether the selected embedding pixels or its neighbors lies at the boundary of the image or not. Data embedding are done by mapping each two or four bits of the secret message in each of the neighbor pixel based on some features of that pixel. Fig.5 and Fig.6 shows the mapping information for embedding two bits or four bits respectively.

PAIR OF MSG BIT	PIXEL INTENSITY VALUE	NO OF ONES (BIN)
01	EVEN	ODD
10	ODD	EVEN
00	EVEN	EVEN
11	ODD	ODD

Fig. 5. Mapping Technique for embedding of two bits

2ND BIT & 3RD BIT/PAIR OF MSG BITS		PIXEL INTENSITY VALUE	NO OF ONES (BIN)
01	00	EVEN	ODD
	01		
	10		
	11		
10	00	ODD	EVEN
	01		
	10		
	11		
00	00	EVEN	EVEN
	01		
	10		
	11		
11	00	ODD	ODD
	01		
	10		
	11		

Fig. 6. Mapping Technique for embedding of four bits

Extraction process starts again by selecting the same pixels required during embedding. At the receiver side other different reverse operations has been carried out to get back the original information.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

IV. ALGORITHMS

Let C be the original 8 bit gray scale image of size N x N i.e. C = $(P_{ij} \mid 0 \leq i < N, 0 \leq j < N, P_{ij} \in 0, 1, \ldots, 255)$. Let MSG be the n bit secret message represented as MSG =$(m_k \mid 0 \leq k < n, m_k \in 0, 1)$.A seed pixel P_{rc} can be selected with row (r) and column (c). Next step is to find the 8 neighbors $P_{r'c'}$ of the pixel P_{rc} such that $r' = r + l$, $c' = c + l$,$-1 \leq l \leq 1$. The embedding process will be finished when all the bits of every bytes of secret message are mapped or embedded.

A. Data Embedding Method for embedding of two bits

Algorithm of the embedding method are described as :

- Input : Cover Image(C), Message (MSG).
- Find the first seed pixel P_{rc}.
- $count = 1$.
- while ($count \leq n$)
- begin (for embedding message in message surrounding a seed pixel).
- cnt=Count number of ones of one of the $P_{r'c'}$ of intensity (V).
- m_k=Get next msg bit.
- $count = count + 1$.
- m_{k+1}=Get next msg bit.
- $count = count + 1$.
- Bincvr= Binary of V.
- If($m_k = 0$ & $m_{k+1} = 1$)
- $Bincvr(zerothbit) = 0$
- If($cnt \bmod 2 = 0$)
- $Bincvr(firstbit) = \neg Bincvr(firstbit)$
- If($m_k = 0$ & $m_{k+1} = 0$)
- $Bincvr(zerothbit) = 1$
- If($cnt \div 2 \neq 0$)
- $Bincvr(firstbit) = \neg Bincvr(firstbit)$
- If($m_k = 0$ & $m_{k+1} = 0$)
- $Bincvr(zerothbit) = 0$
- If($cnt \bmod 2 \neq 0$)
- $Bincvr(firstbit) = \neg Bincvr(firstbit)$
- If($m_k = 0$ & $m_{k+1} = 1$)
- $Bincvr(zerothbit) = 1$
- If($cnt \bmod 2 = 0$)
- $Bincvr(firstbit) = \neg Bincvr(firstbit)$
- End
- Get the next neighbor pixel $P_{r'c'}$ for embedding based on previous $P_{r'c'}$ and repeat.
- End
- Return the stego image (S).

B. Data Extraction Method for extraction of two bits

The process of extraction proceeds by selecting those same pixel with their neighbors. The extracting process will be finished when all the bits of every bytes of secret message are extracted. Algorithm of the extraction method are described as :

- Input : Stego image (S) , count.
- $count = count \div 2$.

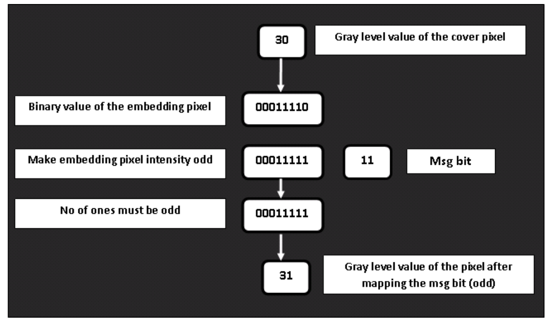


Fig. 7. A snapshot of data embedding process for two bits

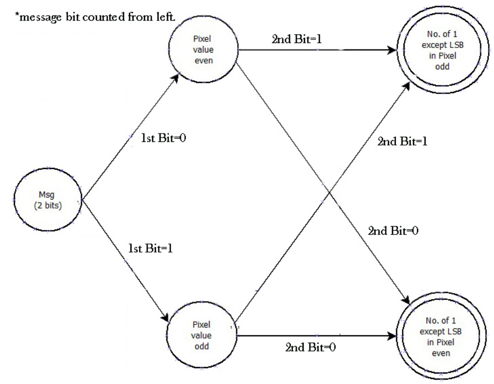


Fig. 8. DFA for embedding process of two bits.

- BinMsg= " ".
- Find the first seed pixel P_{rc}.
- I=0.
- While ($count \leq N$)
- begin (for extract message in message around a seed pixel).
- Get the (First/Next) neighbor pixel $P_{r'c'}$.
- cnt=Count number of ones of one of the $P_{r'c'}$ of intensity (V).
- Bincvr= Binary of V.
- Binmsg(i)=ZerothBit of Bincvr.
- $count = count + 1$.
- $i = i + 1$.
- Binmsg(i)=Enters according to One of ones in the intensity(1 for odd :0 for even).
- $i = i + 1$.
- End.
- Get the next neighbor pixel $P_{r'c'}$ for embedding based on previous $P_{r'c'}$ and repeat.
- End loop.
- Binmsg is converted back to Original message.
- Return Original Message.
- End.

C. Data Embedding Method for embedding four bits

Algorithm of the embedding method are described as :

- Input : Cover Image(C), Message (MSG).
- Find the first seed pixel P_{rc}.
- $count = 1$.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

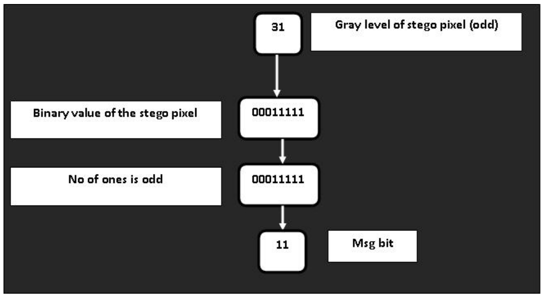


Fig. 9. A snapshot of data extracting process for extraction of two bits

- while ($count \leq n$)
- begin (for embedding message in message surrounding a seed pixel).
- m_k=Get next msg bit.
- $count = count + 1$.
- Mask the 5TH bit from left with the m_k in 'Bincvr'
- m_{k+1}=Get next msg bit.
- $count = count + 1$.
- Mask the 6TH bit from left with the m_{k+1} in 'Bincvr'
- cnt=Count number of ones of one of the $P_{r'c'}$ of intensity (V).
- m_{k+2}=Get next msg bit.
- $count = count + 1$.
- m_{k+3}=Get next msg bit.
- $count = count + 1$.
- Bincvr= Binary of V.
- If($m_{k+2} = 0$ & $m_{k+3} = 1$)
- $Bincvr(zerothbit) = 0$
- If($cnt \bmod 2 = 0$)
- $Bincvr(firstbit) = \neg Bincvr(firstbit)$
- If($m_{k+2} = 0$ & $m_{k+3} = 0$)
- $Bincvr(zerothbit) = 1$
- If($cnt \div 2 \neq 0$)
- $Bincvr(firstbit) = \neg Bincvr(firstbit)$
- If($m_{k+2} = 0$ & $m_{k+3} = 0$)
- $Bincvr(zerothbit) = 0$
- If($cnt \bmod 2 \neq 0$)
- $Bincvr(firstbit) = \neg Bincvr(firstbit)$
- If($m_{k+2} = 0$ & $m_{k+3} = 1$)
- $Bincvr(zerothbit) = 1$
- If($cnt \bmod 2 = 0$)
- $Bincvr(firstbit) = \neg Bincvr(firstbit)$
- End
- Get the next neighbor pixel $P_{r'c'}$ for embedding based on previous $P_{r'c'}$ and repeat.
- End
- Return the stego image (S).

D. Data Extraction Method for extracting four bits

The process of extraction proceeds by selecting those same pixel with their neighbors. The extracting process will be finished when all the bits of every bytes of secret message are extracted. Algorithm of the extraction method are described as :

- Input : Stego image (S) , count.
- $count = count \div 2$.
- BinMsg= " ".
- Find the first seed pixel P_{rc}.
- I=0.
- While ($count \leq N$)
- begin (for extract message in message around a seed pixel).
- Get the (First/Next) neighbor pixel $P_{r'c'}$.
- cnt=Count number of ones of one of the $P_{r'c'}$ of intensity (V).
- Bincvr= Binary of V.
- Binmsg(i)=3rd Bit of Bincvr from Right.
- $i = i + 1$.
- Binmsg(i)=2nd Bit of Bincvr from Right.
- $i = i + 1$.
- Binmsg(i)=ZerothBit of Bincvr.
- $i = i + 1$.
- If (cnt mod 2 = 0) (i.e. it is even) Binmsg(i)=0 Else Binmsg(i)=1
- Binmsg(i)=Enters according to One of ones in the intensity(1 for odd :0 for even).
- $i = i + 1$.
- $count = count + 1$.
- End.
- Get the next neighbor pixel $P_{r'c'}$ for embedding based on previous $P_{r'c'}$ and repeat.
- End loop.
- Binmsg is converted back to Original message.
- Return Original Message.
- End.

One important point needs to be kept in mind that a specific order for selecting the neighbors of the seed pixel has to be maintained for embedding / mapping process and also for the process of extraction other wise it would not be possible to retrieve the data in proper sequence. This sequence has been shown in Figure 8.

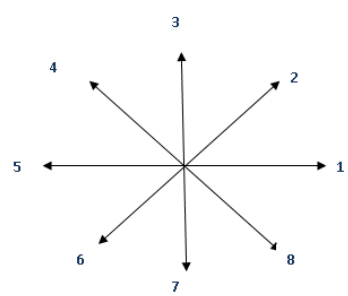


Fig. 10. Sequence of data embedding

E. Pixel Selection Method

Random Pixel Generation for embedding message bits is dependent on the intensity value of the previous pixel selected.It includes a decision factor (dp) which is dependent on intensity with a fixed way of calculating the next pixel.The algorithm for selection of pixel for embedding is described below:

- Input: C , previous pixel position (x,y),pixel intensity value (v).
- Consider dp (Decision Factor)=1 if $(intensity \leq 80)$,dp=2 if $(intensity \geq 80$ & $\leq 160)$,dp=3 if $(intensity > 160$ & $\leq 255)$.
- $t = x + 2 + dp$
- if $(t \geq N) m = 2, n = y + 2 + dp$
- else $m = x + 2 + dp, n = y$
- Return m and n.
- End

122	45	69	132	256	145	56	79	112
156	125	169	123	79	78	12	186	123
224	212	145	125	147	86	45	110	236
119	248	46	112	48	23	79	45	90
119	79	116	189	53	63	130	90	141
56	71	26	83	43	75	93	67	116
90	112	179	212	201	38	99	119	157
83	53	89	115	63	78	90	76	255
131	141	176	159	126	146	255	73	86

Fig. 11. Snapshot of Selected Pixel for embedding.

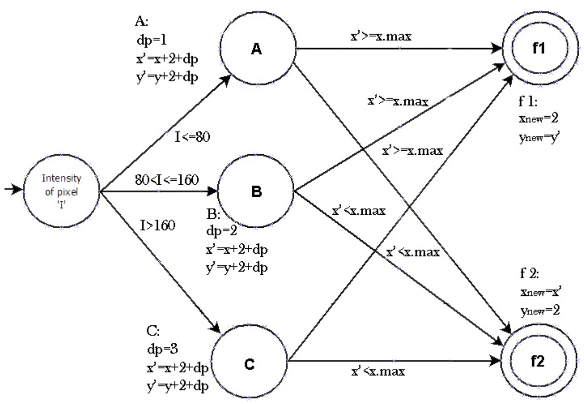


Fig. 12. DFA for pixel selection.

V. Experimental Results

In this section the authors present the experimental results of the proposed method based on two benchmarks techniques to evaluate the data hiding performance based on embedding of two bits or four bits respectively. First one is the capacity of hiding data and another one is the imperceptibility of the stego image, also called the quality of stego image. The quality of stego-image should be acceptable by human eyes. The authors also present a comparative study of the proposed methods with the existing methods like PVD,GLM and the methods proposed by Ahmad T et al.by computing embedding capacity, mean square error (MSE) and peak signal-to noise ratio (PSNR).The authors also compute the normalized cross correlation coefficient for computing the similarity measure between the cover image and stego image. In this section experimental result of stego image are shown based on two well known images: Lena and Pepper. In Fig 13 a segment of Lena as cover image has been shown. Fig 14 shows the same segment of Lena as stego image after embedding the message (two bits per pixel) **"I am an Indian"** on that segment.

41	12	122	34	123	38	64	57	56	89
12	23	74	34	53	75	49	54	67	54
75	87	91	94	97	97	94	95	97	97
96	94	95	98	97	96	97	96	93	95
20	18	24	18	18	15	16	18	13	17
76	68	55	45	29	17	20	19	14	12
90	88	87	88	85	88	88	86	85	87
78	78	79	82	78	74	72	61	64	66
83	84	89	81	81	78	67	55	38	27
91	91	95	90	87	87	90	89	90	93

Fig. 13. A Segment of Cover Image with selected pixel

41	12	120	35	120	36	67	57	59	89
12	23	74	34	53	74	50	54	64	54
74	87	90	93	99	97	94	92	97	97
97	95	94	97	99	97	97	96	93	95
22	18	26	16	18	14	16	18	13	17
76	68	55	44	31	18	20	19	14	12
91	90	84	91	87	88	88	86	85	87
76	78	78	82	78	74	72	61	64	66
80	86	91	80	82	79	67	55	38	27
91	91	95	90	87	87	90	89	90	93

Fig. 14. A Segment of Stego Image with selected pixel with the embedded msg segment "I am an Indian" (two bits per pixel)

In Fig 15 shows the segment of Lena as cover image and Fig 16 shows the same segment of Lena as stego image after embedding the message (four bits per pixel) **"I am an Indian, India is my country"** on that segment.

41	12	122	34	123	38	64	57	56	89
12	23	74	34	53	75	49	54	67	54
75	87	91	94	97	97	94	95	97	97
96	94	95	98	97	96	97	96	93	95
20	18	24	18	18	15	16	18	13	17
76	68	55	45	29	17	20	19	14	12
90	88	87	88	85	88	88	86	85	87
78	78	79	82	78	74	72	61	64	66
83	84	89	81	81	78	67	55	38	27
91	91	95	90	87	87	90	89	90	93

Fig. 15. A Segment of Cover Image with selected pixel

32	17	124	46	116	42	66	55	51	89
7	23	68	35	53	69	49	54	69	54
66	81	88	93	103	98	88	85	110	97
109	81	80	104	103	98	109	101	86	95
21	18		21	18	7	21	18	7	17
64	69	48	36	21	31	22	23	15	12
86	87	88	83	81	80	80	80	80	87
69	78	68	85	78	67	66	61	71	66
93	87	88	88	87	66	74	53	35	27
91	91	95	90	87	87	90	89	90	93

Fig. 16. A Segment of Stego Image with selected pixel with the embedded msg segment "I am an Indian, India is my country" (four bits per pixel)

In Fig 17 shows the image of Lena as cover and also as stego after embedding the message **"I am an Indian and I**

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

feel proud to an Indian."(four bits per pixel). Fig 18 shows the same with Pepper as the image.

Fig. 17. A) Cover Image B) Stego Image of Lena after embedding "I am an Indian and I feel proud to an Indian."

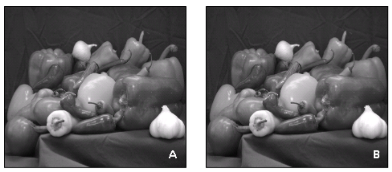


Fig. 18. A) Cover Image B) Stego Image of Pepper after embedding "I am an Indian and I feel proud to an Indian."

A comparative study of the embedding capacity with other methods has been illustrated in figure 19 (two bits per pixel) and figure 20 (four bits per pixel) respectively.

IMAGE	IMAGE SIZE	PVD	GLM	AHMAD ET ALL.	PMM
LENA	128x128	**	2048	2493	2393
	256x256	**	8192	10007	10012
	512x512	50960	32768	40017	45340
PEPPER	128x128	**	2048	2443	2860
	256x256	**	8192	9767	11694
	512x512	50685	32768	39034	46592

Fig. 19. Comparision of embedding capacity for two bits

IMAGE	IMAGE SIZE	PVD	GLM	AHMAD ET ALL.	PMM
LENA	128x128	**	2048	2493	4786
	256x256	**	8192	10007	20024
	512x512	50960	32768	40017	90680
PEPPER	128x128	**	2048	2443	5720
	256x256	**	8192	9767	23388
	512x512	50685	32768	39034	93184

Fig. 20. Comparision of embedding capacity for four bits

** For PVD method all the images used are of size 512x512.

A. Peak Signal to Noise Ratio (PSNR)

PSNR measures the quality of the image by comparing the original image or cover image with the stego-image, i.e. it measures the percentage of the stego data to the image percentage. The PSNR is used to evaluate the quality of the stego-image after embedding the secret message in the cover. Assume a cover image C(i,j) that contains N by N pixels and a stego image S(i,j) where S is generated by embedding / mapping the message bit stream. Mean squared error (MSE) of the stego image as follows:

$$MSE = \frac{1}{[N \times N]} \sum_{i=1}^{N} \sum_{j=1}^{N} [C(ij) - S(ij)]^2$$

The PSNR is computed using the following formulae:

$$PSNR = 10 \log_{10} 255^2 / \ MSE \ db.$$

A comparative study of PSNR of various methods has been illustrated in figure 21 and figure 22 respectively.

IMAGE	IMAGE SIZE	PVD	GLM	AHMAD ET ALL.	PMM
LENA	128x128	36.20	30.5	44.30	49.0296
	256x256	35.00	33.20	46.80	50.3489
	512x512	41.79	35.50	55.00	54.1515
PEPPER	128x128	38.70	38.00	43.50	47.9468
	256x256	35.00	37.20	47.50	48.3668
	512x512	40.97	34.00	52.50	54.1521

Fig. 21. Comparison of PSNR after embedding two bits per pixel

IMAGE	IMAGE SIZE	PVD	GLM	AHMAD ET ALL.	PMM
LENA	128x128	36.20	30.5	44.30	36.5864
	256x256	35.00	33.20	46.80	36.0547
	512x512	41.79	35.50	55.00	34.7396
PEPPER	128x128	38.70	38.00	43.50	34.9404
	256x256	35.00	37.20	47.50	36.2118
	512x512	40.97	34.00	52.50	36.9247

Fig. 22. Comparison of PSNR after embedding four bits per pixel

B. Similarity Measure

For comparing the similarity between cover image and the stego image, the normalized cross correlation coefficient (r) has been computed. In statistics, correlation indicates the strength and direction of a linear relationship between two random variables. The correlation coefficient ρ_{xy} between two random variables X and Y with expected values μ_x andμ_y and standard deviations σ_x and σ_y is defined as

$$\rho_{x,y} = \frac{cov(x,y)}{\sigma_x \sigma_y} = \frac{E((X-\mu_x)(Y-\mu_y))}{\sigma_x \sigma_y}$$

where E is the expected value operator and cov means covariance. The value of correlation is 1 in the case of an increasing linear relationship, -1 in the case of a decreasing linear relationship, and some value in between in all other cases, indicating the degree of linear dependence between the variables.

Cross correlation is a standard method of estimating the degree to which two series are correlated. Consider two series x(i) and y(i) where i=0,1,2,. . . ,N-1. The cross correlation r at delay d is defined as

$$r = \frac{\sum_i [(x(i) - mx)(y(i-d) - my)]}{\sqrt{\sum_i (x(i) - mx)^2} \sqrt{\sum_i (y(i-d) - my)^2}}$$

where mx and my are the means of the corresponding series. The cross-correlation is used for template matching which is motivated through the following formula

$$r = \sum_{\substack{x \\ y}} f(x,y) t(x-u, y-v)$$

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

where f is the image and the sum is over x, y under the window containing the feature t positioned at u, v.

Similarity measure of two images can be done with the help of normalized cross correlation generated from the above concept using the following formula:

$$r = \frac{\sum_{(C(i,j)-m_1)(S(i,j)-m_2)}}{\sqrt{(\sum_{C(i,j)-m_1})^2}\sqrt{(\sum_{S(i,j)-m_2})^2}}$$

Here C is the cover image, S is the stego image,m_1 is the mean pixel value of the cover image and m_2 is the mean pixel value of stego image. It has been seen that the correlation coefficient computed here for all the images is almost one which indicates the both the cover image and stego image are of highly correlated i.e. both of these two images are same.

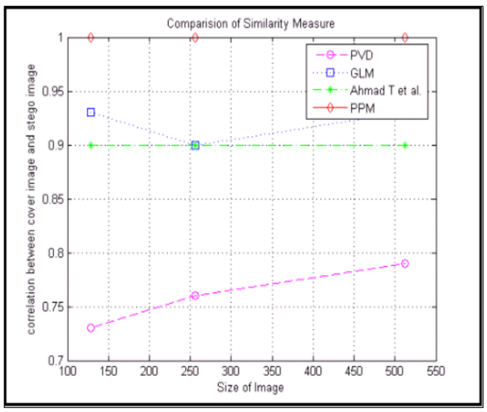


Fig. 23. Comparision of Similarity Measure for Lena

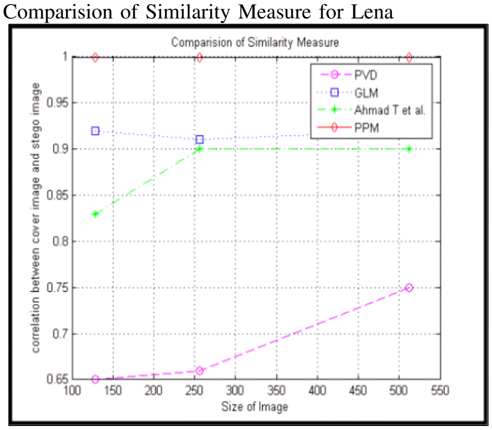


Fig. 24. Comparision of Similarity Measure for Pepper

VI. Analysis of the Results

In this article the authors proposed an efficient image based steganography approach for hiding information in a gray scale image. Comparison has been shown with some existing methods like PVD, GLM and the technique proposed by Ahmad T et al. From the experimental results in can be seen that the embedding capacity of the proposed method is better compared to PVD, GLM and the other technique in most cases and also the similarity measures proves that the proposed method better among these four methods which ensures that cover image and the stego image is almost identical. As the message bits are not directly embedded at the pixels of the cover image, steganalysis may be able to find out the embedded bits but can not be able to extract the original message bits.PSNR value of the proposed method (two bits per pixel) for various sizes of the image better than compared to other methods.

VII. Conclusion

The work dealt with the techniques for steganography as related to gray scale image. A new and efficient steganographic method for embedding secret messages into images without producing any major changes has been proposed. Although in this method it has been shown that each two bit or four bit of the secret message has been mapped in the pixels of the cover image,but this method can be extended to map 8 no of bits per pixel by considering more no of features of the embedding pixels.This method also capable of extracting the secret message without the cover image. This approach may be modified to work on color images also.

References

[1] Ahmad T. Al-Taani. and Abdullah M. AL-Issa. A novel steganographic method for gray-level images. *International Journal of Computer, Information, and Systems Science, and Engineering*, 3, 2009.

[2] RJ Anderson. Stretching the limits of steganography. *Information Hiding, Springer Lecture Notes in Computer Science*, 1174:39–48, 1996.

[3] Ross J. Anderson. and Fabien A.P.Petitcolas. On the limits of steganography. *IEEE Journal on Selected Areas in Communications (J-SAC), Special Issue on Copyright and Privacy Protection*, 16:474–481, 1998.

[4] Souvik Bhattacharyya. and Gautam Sanyal. Study of secure steganography model. In *Proceedings of International Conference on AdvancedComputing and Communication Technologies (ICACCT-2008)*, Panipath,India, 2008.

[5] Souvik Bhattacharyya. and Gautam Sanyal. An image based steganography model for promoting global cyber security. In *Proceedings of International Conference on Systemics, Cybernetics and Informatics*, Hyderabad,India, 2009.

[6] Souvik Bhattacharyya. and Gautam Sanyal. Implementation and design of an image based steganographic model. In *Proceedings of IEEE International Advance Computing Conference*, Patiala ,India, 2009.

[7] Souvik Bhattacharyya. and Gautam Sanyal. Hiding data in images using pixel mapping method (pmm). In *Proceedings of 9th annual Conference on Security and Management (SAM) under The 2010 World Congress in Computer Science,Computer Engineering, and Applied Computing(WorldComp 2010)*, LasVegas,USA, July 12-15,2010.

[8] J.Y. Hsiao. C.C. Chang. and C.-S. Chan. Finding optimal least-significant-bit substitution in image hiding by dynamic programming strategy. *Pattern Recognition*, 36:1583–1595, 2003.

[9] C.K. Chan. and L. M.Cheng. Hiding data in images by simple lsb substitution. *Pattern Recognition*, 37:469–474, 2004.

[10] Scott. Craver. On public-key steganography in the presence of an active warden. In *Proceedings of 2nd International Workshop on Information Hiding.*, pages 355–368, Portland,Oregon, USA, 1998.

[11] Potdar V.and Chang E. Gray level modification steganography for secret communication. In *IEEE International Conference on Industria lInformatics.*, pages 355–368, Berlin, Germany, 2004.

[12] P Huang. K.C. Chang., C.P Chang. and T.M Tu. A novel image steganography method using tri-way pixel value differencing. *Journal of Multimedia*, 3, 2008.

[13] Y. K. Lee. and L. H.Chen. High capacity image steganographic model. *IEE Proc.-Vision, Image and Signal Processing*, 147:288–294, 2000.

[14] N.F.Johnson. and S. Jajodia. Steganography: seeing the unseen. *IEEE Computer*, 16:26–34, 1998.

[15] C.F. Lin. R.Z. Wang. and J.C. Lin. Image hiding by optimal lsb substitution and genetic algorithm. *Pattern Recognition*, 34:671–683, 2001.

[16] Gustavus J. Simmons. The prisoners' problem and the subliminal channel. *Proceedings of CRYPTO.*, 83:51–67, 1984.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[17] JHP Eloff. T Mrkel. and MS Olivier. An overview of image steganography. In *Proceedings of the fifth annual Information Security South Africa Conference.*, 2005.
[18] D.C. Wu. and W.H. Tsai. A steganographic method for images by pixel-value differencing. *Pattern Recognition Letters*, 24:1613–1626, 2003.

Souvik Bhattacharyya received his B.E. degree in Computer Science and Technology from B.E. College, Shibpur, India, presently known as Bengal Engineering and Science University (BESU) and M.Tech degree in Computer Science and Engineering from National Institute of Technology, Durgapur, India. Currently he is working as a Senior Lecturer in Computer Science and Engineering Department at University Institute of Technology, The University of Burdwan. He has a good no of research publication in his credit. His areas of interest are Natural Language Processing, Network Security and Image Processing.

Dr. Lalan Kumar received his Ph.D. degree from the Indian School of Mines(ISM), Dhanbad Jharkhand. Joined National Informatics (NIC) Centre, under Planning Commission of Govt. of India in 1990 and worked till 25th Nov.'02. Joined Central Institute of Mining and Fuel Research (CIMFR) on 25th Nov.'02. Prior to joining CMRI as Scientist, he has studied, designed, developed and implemented many packages for the District, state and some of the packages are running in almost all the districts of the country. He has been appointed as a panel expert for local governance and community engagement for the various departments of state government. He has published more than 50 papers in International and National Journals of repute. He is member of many advisory board/Review committee/Chairman/Resource person of Universities/journals/ International/national Seminar cum Symposia/Institutions.Dr.Kumar has organized many International and National seminar cum exhibition time to time and edited books.

Gautam Sanyal has received his B.E and M.Tech degree from Regional Engineering College (REC), Durgapur, now, National Institute of Technology (NIT), Durgapur, West Bengal, India. He has received Ph.D (Engg.) from Jadavpur University, Kolkata, West Bengal, India, in the area of Robot Vision. He possesses an experience of more than 25 years in the field of teaching and research. He has published nearly 40 research papers in International and National Journals / Conferences. His current research interests include Natural Language Processing, Stochastic modeling of network traffic, High Performance Computing, Computer Vision. He is presently working as a Professor in the department of Computer Science and Engineering and also holding the post of Dean (Student's Welfare) at National Institute of Technology, Durgapur, West Bengal, India.

Matching SHIQ Ontologies

B.O. Akinkunmi, A.O. Osofisan, and A.F. Donfack Kana
ope34648@yahoo.com ,nikeosofisan@gmail.com, donfackkana@yahoo.fr
Department of Computer Science,
University of Ibadan, Nigeria.

Abstract- *This paper proposes and evaluates an approach to facilitate semantic interoperability between Ontologies built in SHIQ description logic language in an attempt to overcome the heterogeneity problem of Ontologies. The structural definition of Ontologies is used as a key point to predict their similarities. Based on SHIQ General Concept Inclusion, Ontologies to be mapped are translated into hierarchical trees and a graph matching technique is used to find out similarities between the trees. Similarity between concepts is predicted based on their level of hierarchy and their logical definition. Semantic similarities between concepts are evaluated by putting more emphasis on the logical operators used in defining concepts with less reference to concepts syntactic similarities analysis. The obtained result shows that a pure structural comparison based mainly on logical operators used in defining Ontologies concepts provides a better approximation than a comparison combining the logical and syntactic similarities analysis evaluated based on the edit distance function.*

Keywords: *Ontology, Description logics, Mapping, Interoperability, Semantic.*

1. Introduction

Due to heterogeneity of information sources, the need has arisen for the development of techniques for representing information in an unambiguous way and finding approaches that allow applications to simultaneously manipulate data available in multiple sources. Ontologies [15] have been used to provide structured knowledge and information that give a common understanding of the domain been modeled.

The non standardization of Ontologies and Ontologies languages leads to the existence of several incompatible Ontologies systems even those representing the same domain. This means that, different ontology builders may have different approaches to model a given domain or can use different terminology to represent the same concept, which may be semantically the same but syntactically different. This leads to inconsistent interpretation and incompatibility between Ontologies even those describing the same domain. Incompatibility limits the sharing of knowledge between machines or humans who do not have the same understanding of what terminology means. One cannot talk of common understanding of domains as long as there is inconsistent interpretation of terminologies. Consequently, interoperability among Ontologies systems is no longer possible at semantic level unless there is a mean of matching Ontologies created by different builders. Ontologies then face the same or even harder problems with respect to heterogeneity as any other piece of information [16]. Several research works have been undertaking in recent time in this field to determine how Ontologies should be constructed and to attempts to overcome the problem of semantic interoperability. Description logics(DL) [1,12] are becoming widely used in Ontologies building, as a language that provides a clear semantic and inference services to support the design, evolution, maintenance and integration of Ontologies. Despite the merit of description logics as a language for Ontologies building, interoperability among description logics based Ontologies is still an unsolved problem. Several authors e.g. [2,3] show how description logics can be used as Ontologies language but there have been few effective and efficient proposals on how interoperability among them can really be achieved.

An approach to this problem is to determine a platform that reconciles different Ontologies. Literatures on how to reconcile different Ontologies remain scanty [17,11]. Mapping of Ontologies elements is often used as an attempt for that purpose. A review of existing systems for ontology mapping reveal that most systems evaluate the semantic similarities by performing a deep analysis of syntactic comparison or by using class instance comparison[4,9,13]. The main drawback of this approach is that, in ontology building, some jargon may not be easily interpreted so as to derive their meaning from their syntax. Since most natural languages are not yet standardized, syntactic analysis cannot be seen as a perfect solution.

In this paper, we describe a mapping approach that can be used to achieve interoperability between SHIQ Ontologies by performing a matching between Ontologies based mostly on the hierarchical structure of Ontologies and constructors used in defining concepts. Textual Ontologies are translated into hierarchical trees based on SHIQ general concept inclusion and graph matching techniques are used to identify similar concepts. Similarities between concepts are evaluated based on their logical definition, the similarities of their ancestors and that of their successors.

In the next section we present a formal definition of ontology mapping. Section 3 introduces SHIQ language and it syntax, section 4 describes the proposed SHIQ mapping system and section 5 focuses on the implementation and the discussion of the obtained result

2. Mapping Ontologies

Ontology mapping is the main task required to achieve interoperability between two agents or services using different Ontologies. Mapping is a set of formulae that provide the semantic relationship between concepts in the models [10]. Mapping is to establish correspondences among different Ontologies and to determine the set of overlapping concepts (concepts that are similar in meaning but have different name or structure) and concepts that are unique to each of the sources [14]. Given two Ontologies O_1 and O_2,

mapping Ontologies O_1 onto ontology O_2 means that for each entity in ontology O_1, we try to find a corresponding entity, which has the same intended meaning, in ontology O_2.

Mapping takes as input two lists of terms from Ontologies O_1 and O_2 and produces a list of matched pairs. Each pair contains two terms: one from the source ontology O_1 and another from the target ontology O_2. Ontology mapping as a task will entail discovering the following properties:

Property 2.1 Equality:

Two Ontologies are equal if there exist a mapping that can transform ontology O_1 into ontology O_2

Property 2.2 Subsumption:

Ontology O_1 subsumes O_2 if a mapping exists that can transform a subset O_1' of O_1 into O_2.

Property 2.3 Intersection:

Intersection occurs between ontology O_1 and O_2 if they have some overlapping area. That is there exist O_1' and O_2' such that $O_1' \subseteq O_1$ and $O_2' \subseteq O_2$ and a mapping exist that can transform O_1' into O_2'.

Property 2.4 Disjointeness:

Ontology O_1 and O_2 are disjoint when the two Ontologies are totally different.

Ontology is made up of a collection of related concepts. Analyzing the mapping properties between two different Ontologies is the task of finding the mapping concepts in the two Ontologies. This allows us to characterize relationship between Ontologies as relation between their concepts. Map$\subseteq C_1$ x C_2 where C_1 and C_2 are sets of concepts in ontology O_1 and O_2 respectively is regarded as one to one function between Ontologies concepts. Each concept of a given ontology may be associated with a corresponding semantically similar concept in the other ontology.

3. Review of SHIQ description logics as ontology language.

Description logics provide languages for the description of concepts, relations used to represent knowledge. They can be used in ontology building. They are characterized by the use of various constructors to build complex concepts from simpler ones, with an emphasis on the decidability as key reasoning tasks, and by the provision of sound, complete and (empirically) tractable reasoning services [5].

In contrast to most description logic which concentrates on constructors to describe concepts, SHIQ DL[5,6,7,8] is *a*n expressive knowledge representation formalism that extends *ALC* description logics with qualifying number restrictions, inverse roles, role inclusion axioms (RIAs) R S, and transitive roles[6].

3.1. Syntax of SHIQ concepts

Concepts are used to describe the relevant notions of an application domain. The terminology (TBox) introduces abbreviations (names) for complex concepts. In SHIQ, the TBox allows one to state also more complex constraints. Let N_C be a set of concept names. The set of SHIQ-concepts is the smallest set such that the following hold:

1. Every concept name A $\in$ N_C is a SHIQ-concept.
2. If C and D are SHIQ-concepts and r is a SHIQ-role, then C D, C D, $\neg$C, $\forall$r.C, and $\exists$r.C are SHIQ-concepts,
3. If C is a SHIQ-concept, r is a simple SHIQ-role, and n$\in$**N**, then ($\leq$nr.C) and ($\geq$nr.C) are SHIQ-concepts.

A general concept inclusion (GCI) is of the form C D, where C and D are SHIQ-concepts. A finite set of GCIs is called a TBox. Concept definition is of the form A C, where A is a concept name. It can be seen as an abbreviation for the two GCIs A C and C A. For example, a subset of the definition of a family can be stated in Tbox using GCI as follow:

man human $\geq$haschild.human
woman human female
man human $\neg$ woman
mother woman $\geq$1haschild.human
father man $\geq$1haschild.human
grandmother mother $\exists$ haschild.parent
grandfather father $\exists$haschild.parent
motherwith5children mother $\geq$5haschild.human $\leq$2haschild. human
motherwithoutdaughter mother $\forall$ haschild.man

3.2. Syntax of SHIQ-roles

Role is interpreted as binary relation. It uses binary relations (such as Boolean operators, composition, inverse, and transitive closure) as role forming constructors. Syntax of SHIQ role is defined as follows:

1 Every role name is a role description (atomic role), and if *R*, *S* are role descriptions, then R, S, *R* *S* (intersection), *R* *S* (union), are normal role.
For example, one can express that someone has a number of male children within the range of 2 and 7 by ($\geq$2 has_Child $\leq$ 7 has_Child).male

2 Let R be a set of role names, which is partitioned into a set R_+ of transitive roles and a set R_P of normal roles. The set of all SHIQ-roles is R {r^- / r R}, where r^- is called the inverse of the role r.

The inverse role construct allows one to denote the inverse of a given relation. One can for example one can state with has_child$^-$.doctor that someone has a parent who is a doctor, by making use of the inverse of role has_child. The inverse relation on binary relations is symmetric, i.e., the inverse of r^- is again r.

A role inclusion axiom is of the form r s where r and s are SHIQ-roles. A role hierarchy is a finite set of role inclusion axioms. Role inclusion is transitive and an inclusion relation between two role transfers to their inverses

3.3. Describing Ontologies in SHIQ

In general, ontology can be formalized in a TBox as follows:

Firstly, we restrict the possible worlds by introducing restrictions on the allowed interpretations using the general concepts inclusion (GCIs).

Secondly, we can define the relevant notions of our application domain using concept definitions. A concept name is called defined if it occurs on the left-hand side of a definition, and primitive otherwise.

We want our concept definitions to have definitional impact, i.e., the interpretation of the primitive concept and role names should uniquely determine the interpretation of the defined concept names. For this, the set of concept definitions together with the additional GCIs must satisfy three conditions:

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

- There are no multiple definitions, i.e., each defined concept name must occur at most once as a left-hand side of a concept definition.
- There are no cyclic definitions, i.e., there is no cyclic dependency between the defined names in the set of concept definitions.
- The defined names do not occur in any of the additional GCIs.

4. The SHIQ match system.

The SHIQ match system takes as input two Ontologies built in SHIQ and produce as output the proposed mapped concept from the two Ontologies.

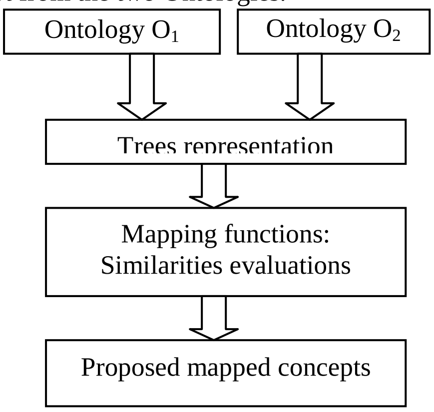


Fig.1 structure of the SHIQ match system

Textual Ontologies are first translated into trees. Trees are traversed and similarities are computed between trees elements based on some metrics to be defined. Finally, the system proposes to the users the matching between concepts of the two Ontologies based on their obtained similarity values

4.1. Parsing Ontologies into trees

Parsing Ontology into tree takes, for each valid general concept inclusion, its constituent parts and adds them into a tree. This represent ontology as a tree T by a couple $T=(N,E)$ where N is a finite set of nodes and $E \subseteq N \times N$ is a set of directed edges. Nodes of the tree represent concept names while edges occurring between two nodes represent the role as well as associated constructors, which is the relation that links the two concepts. In other words, a sub-concept becomes a child of the parent concept from which it was defined, while the node linking them is labeled with associated constructors that define the sub-concept in terms of the parent.

Ontologies to be mapped are defined in negation normal form (NNF), that is, negation occurs only in front of concept names. Any SHIQ concept can easily be transformed to an equivalent one in NNF by pushing negations inwards by using a combination of De Morgan's laws:

(B C)= ¬B ¬C

¬(B C)= ¬B ¬C

and the following equivalences:

$\neg(\exists r.C) = (\forall r.\neg C)$

$\neg(\forall r.C) = (\exists r.\neg C)$

$\neg(\leq n\ r.C) = \geq(n+1)\ r.C)$

$\neg(\geq n\ r.C) = \leq(n-1)\ r.C$

In NNF form, concepts defined using GCI is represented as follow:

$Concept_{def}$ conce pt_{sub} $role_1.concept_1$ $role_2.concept_2$ … $role_n.concept_n$

Where n is a positive integer, which may be zero and $role_i$, used as a generic term that may represent constructors as well as role name.

The concept definition in this case can be divided into three differents parts: The defined concept specified by $Concept_{def}$, which is the concept being defined. It is regarded as a subset of the parent concept. The parent (or subsuming) concept represented by $concept_{sub}$, which is a concept more general than the defined concept and is assumed to have the defined concept as one of its subset. Finally, a sequence of $role_i.concept_i$ separated by a set operator ( ). Each $role_i.concept_i$ term represents a specialization of the parent concept. That is, they specify an attribute of the defined concept that makes it a special kind of the parent concept.

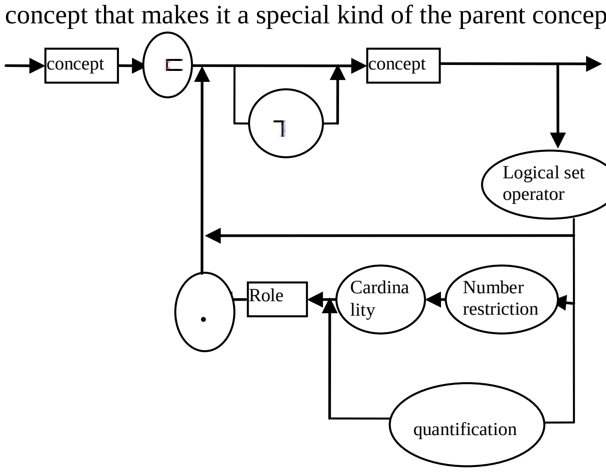


Fig. 2 SHIQ General Concept Inclusion syntax diagram.

Starting from the top concept as a root of the tree, each GCI line is scanned using the GCI diagram and the concept been defined is added in the tree as a successor node of it parent concept. Any other element of the definition (roles, constructors and any associated concepts) appear at the edges linking the two nodes. Note that the parent concept appearing in the right side of the concept inclusion symbol of each definition. This preserves the hierarchical structure of the tree, where concepts of higher level are assumed to subsume concepts of lower level.

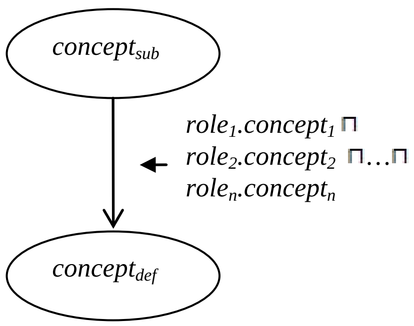


Fig.3 Tree representation of General Concept Inclusion

4.2 Match SHIQ Ontologies from trees.

Comparison between the two trees is performed by traversing the two trees simultaneously and evaluating

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

similarities between trees elements. Normally, each node of the first tree is compared with all the node of the second tree, to find its similar node from all the nodes of the second tree. We have adopted an approach that optimizes the total number of comparison by reducing the searching space by a factor at which concepts are similar.

Nodes to be compared next are selected based on the similarities of their ancestors. When the current selected node of the first tree is compared with all the nodes of the second tree, a decision about its similarity with a node of the second tree is made. Whenever a node of the second tree is perfectly similar with that of the currently selected node of the first tree, the next step of comparison takes only their sub-trees, taking the two identical nodes as roots. For instance, in *fig. 7*, with the matching between *woman* and *femme*, all nodes derived from *woman* should be compared only with the nodes derived from *femme* rather than comparing them with all the nodes of the second tree. In case there is no perfect match, all its successors are compared with all the nodes of the second tree

```
Comparetreee(tree1,tree2: tree)
Begin
    Select a node of tree1 say node1
    While (there still exist unvisited node in tree2)do
    Begin
        Select a node in tree2 say nodei
        Mark nodei as visited
        Compute_ similarity( node1, nodei)
    End
    If (there exist a node of tree2 say nodei where
    computed similarity=1) then
    Begin
        For (all subtree of node1 )do
            For (all subtree of nodei)do
              Comparetree(subtree of node1,
            subtree of nodei)
            End
        End
    End
    Else
    Begin
        For (all subtree of node1 )do
            Comparetree(subtree of node1,
        tree2)
        End
    End
End
```

Pseudo code 1: Algorithm for traversing trees to perform the comparison.

It is worth nothing that *compute_similarity* is a pseudo method that refers the three methods for computing the similarity between nodes. That is comparedefinition, *Comparesuccessor and* importing the parent similarity as stated in fig. 4.

This comparison approach reduces the number of comparison to be performed from the Cartesian product of all the nodes of the two trees (N x M) to a factor of the numbers of successors nodes of the two concepts being compared. Any pair of compared nodes is treated as a candidate mapping that will be checked and the most similar concepts (if there exist) based on the similarity metric to be defined, is considered as the mapped concepts.

Similarity between two concepts is defined as follows:

$$Sim : \mathcal{E}_1 \times \mathcal{E}_2 \times O_1 \times O_2 \in [0, 1]$$

Where $\mathcal{E}_1$, $\mathcal{E}_2$ are the set of concepts of ontology O_1 and O_2 respectively.

Sim function returns a degree of similarity between pairs of concepts. The similarity value is between 0 and 1, where 1 stands for perfect match and 0 for no match.

The hierarchy structure of the tree implies that each concept was defined only once and then appears only once on the tree. Thus each concept of the first ontology may only have at most only one similar concept in the second ontology. With this assumption, the task of mapping between two Ontologies is reduced to a one to one mapping between their respective concepts.

To have a good understanding of the semantic of a particular defined concept, there is a need to understand the semantic of all other concepts associated in its definition. For example, with the following representation:

$$Mother \sqsubseteq Woman \sqcap \geq 1 haschild.Human$$

mother is defined as a *woman* having a child which is an instance of a class *human*. To capture the meaning of the concept *mother*, we need to understand first the meaning of concept *woman* and *human*. SHIQ Ontology is defined using concept name, role name, which is a relation occurring between concepts and possibly constraints on this relation, as well as constructors to combine those concepts. For each pair of concept been compared, a logical comparison is performed on the logical symbols that associate the concepts to their neighbor. This comparison is associated a weight which define the importance of the logical operators in determining the similarities of the concepts been defined. Because theses operators do not have any relation with the natural language used, they are considered to be the main point at which the comparison is based upon. Their comparison is based on a direct logical comparison. Our mapping approach is to make the mapping as free as possible from the syntactic comparison. We assign a high weight to each logical comparison.(fig.5)

Whenever there is a need, generally when Ontologies are built in the same natural language, syntactic comparison may be used to support the logical comparison but with a very low weight. The syntactic comparison may be performed on Meta data used as concepts name or role label. Generally, when two syntactically similar terms are used in the same context, they have the same meaning. It is observed that, in modeling the same world using the same natural language, it will be difficult to find that knowledge modelers have used totally different vocabulary to define all of their concepts and role. In some cases, they may used highly syntactically related to represent the same things. As example, one may

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

decide to use "studentname" to represent the student name while another builder may use "student_name" for the same purpose. The edit distance *(ed)* introduced by Levenshtein is used as a basic measure for the syntactic similarity function. The idea behind the edit distance is to measure the minimum number of token insertions, deletions, and substitutions required to transform one string into another. The difference between two strings can be obtained from the longest common subsequence (LCS) by subtracting the double of the length of the LCS from the sum of the length of the two strings.

$$Ed(string1,string2) = |string1|+|string2|-2(|LCS(string1,string2)|)$$

Based on the edit distance, we define the syntactic similarity function as a maximum value between 0 and the ratio of the minimum value between the lengths of the two strings minus their edit distance over the minimum length of the two strings.

$$Sim_{syntax}(string1,string2)= \max\left(0, \frac{\min(|string1|,|string2|)-ed(string1,string2)}{\min(|string1|,|string2|)}\right)$$

Sim_{syntax} function returns a degree of similarity of two strings, which is a value between 0 and 1, where 1 stands for perfect match and zero for bad match. It considers the number of changes that must be made to transform one string into the other and weights the number of these changes against the length of the shorter of the two strings.

The similarity between concepts is captured based on three major comparisons: Concepts definition, concepts successors and parents' concepts.

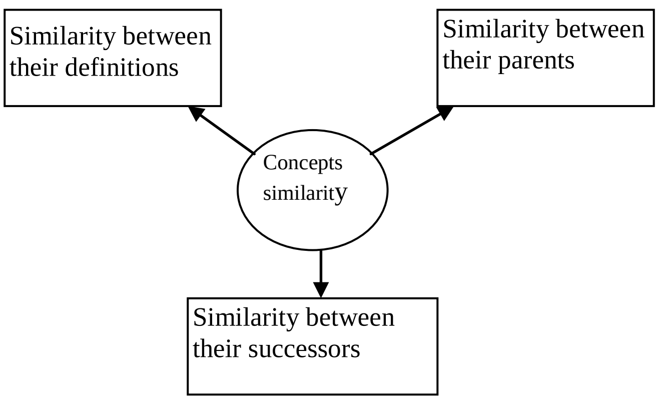


Fig 4. Approach for concepts similarities evaluation

4.3 Definition comparison

Comparing definitions means that we find out similarities between operators, relations as well as other concepts used in the definitions. In a tree, it implies a comparison between the two edges leading to the nodes being compared. A comparison of all the $role_i.concept_i$ of the two definitions is performed. A best match for each $role_i.concept_i$ of the first tree is determined and an average value of their similarity is obtained as the similarity value of the concept definition. The comparison is achieved by performing a syntactic comparison on the role name, a logical comparison on the set operators used in the definition. Generally, $concept_i$ appearing in the $role_i.concept_i$ is an already defined concept. Their similarity value can be obtained just by referring to their already computed value. The pseudo code below describes the comparison between two definitions.

Comparedefinition($node_1$,$node_2$)
Var
Resultdefinition: array of record
Role1:string
Role2:string
Score:real
Begin
  For i=1 to number of $role_i.concept_i$ appearing in the edge leading to node1 do
  begin
  For j=1 to number of $role_j.concept_j$ appearing in the edge leading to $node_2$ do
  Begin
    Scorerolename ←simsyntax($rolename_i$, $rolename_j$)
    scorecardinality←simlogic($cardinality_i$,$cardinality_j$)
    scorequantification←simlogic($quantification_i$, $quantification_j$)
    scorenumberrestriction←simlogic(number $restriction_i$, number $restriction_j$)
    scoreassociatedconcept←similarity value of ($concept_i$,$concept_j$)
    resultdefinition[k].role1←rolei.$concept_i$
    resultdefinition[k].role2←rolej.$concept_j$
    *resultdefinition[k].score←sum(scorerolename*weight +scorecardinality*weight+scorequantification*weight+ corenumberrestriction*weight+ scoreassociatedconcept*weight) /sum(weight)*
    k←k+1
  end
end

for (each $role_i.concept_i$ in resultdefinition[i].role1) do
begin
    select their best match score in resultdefinition[i].score
    averagescore← sum(all best matches)/(number of $role_i.concept_i$ appearing in the edge that lead to $node_1$)
    scoreconceptdefinition←averagescore
    return scoreconceptdefinition
  end
end

Pseudo code2: Algorithm for comparing the definition of two concepts

4.4 Successor's comparison

It is evident that similar concept will also subsume similar children concepts. So, comparison is performed on all the concepts definitions that use each of the two concepts as a parent concept. To achieve this, we compare all the edges departing from both nodes being compared. For two given nodes, we perform a Cartesian product of the number of children of each node. For instance, in *fig.7*, if the two nodes being compared are *mother* and *maman*, we find also the similarities between the sets *(grandmere, merede5enfants)* and *(grandmother, motherwith5children)* as they are

successor's node of *maman* and *mother* respectively. For each edge departing from the node being compared in the first tree, it best match is detected from all the edges of the node of the second tree. An average value is obtained from the ratio of all the best match value and the number of edges departing from the node of the first tree.

Comparesuccessor($node_1$,$node_2$)
Var
Resultsuccessor: array of record
concept1:string
concept2:string
Score:real
Begin
For i=1 to number of successor of node1 do
Begin
For j=1 to number of successor of node2 do
Begin
resultsuccessor [k].concept1← $successor_i$
resultsuccessor [k]. concept2← $successor_j$
resultsuccessor[k].score←conceptdefinition(successor$_i$, $successor_j$)
k←k+1
end
end
for (each successor$_i$ *in resultsuccessor[i].concept1) do*
begin
select their best match score in resultsuccessor[i].score
averagescore← sum(all best matches)/(number of successor of node1)
scoresuccessor←averagescore
return scoresuccessor
end
end

Pseudo code3: Algorithm for computing the similarity between successors' nodes

4.5 Parent node comparison

In a real world, two things are similar if they produce the same elements. Since Ontology is just the modeling of the real world, if two concepts are similar, then there should be a similarity between their parents. For each pair of concepts being compared, we refer to the similarity of their parents to predict their own similarity.

With the comparison been evaluated at three different level, we obtain, three values for a candidate pair of concepts from the two Ontologies. We define an aggregate value that gives the similarity value for that mapping. These values are assigned different weights that represent the importance for each of those above obtained value, to influence the final similarity value. We define the result of similarity between two concepts as follow:

Weight parent(score parent)+ weight concept definition(score concept definition)+weight concepts subsumption(score concept subsumption) /Σ(weight)

A syntactic comparison (see *fig.5*), should be avoided whenever the two Ontologies are not in the same natural language. The figure 5 below shows all the comparison performed for a given pair of concept to deduce their similarities

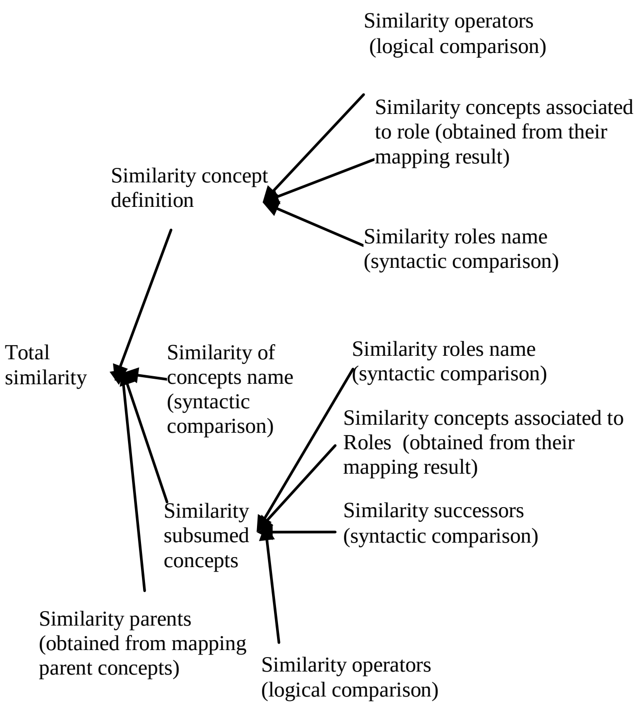


Fig 5: Summary of comparisons.

5. Implementation and result

SHIQ match system consists of a user interface and an associated mapping engine built based on the above described technique of mapping. It provides facilities of building Ontologies to be compared. The system emphasis on the internal matching structure rather than the visualization aspect of transformations occurred on the ontology.

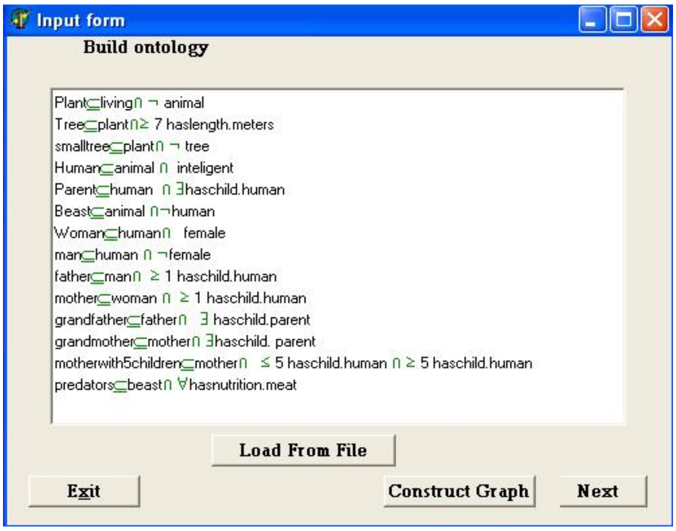


Fig.6 the SHIQ match system interface for Ontology building

Taking as example, the two given Ontologies, the first describing a subset of a living things and the second ontology, it subset describing the family bellow:

Ontology1

Animal $\sqsubseteq$ living
Plant $\sqsubseteq$ living $\sqcap \neg$ animal
Tree $\sqsubseteq$ plant $\sqcap >7$ haslength.meters
smalltree $\sqsubseteq$ plant $\sqcap \neg$ tree
Human $\sqsubseteq$ animal $\sqcap$ intelligent
Parent $\sqsubseteq$ human $\sqcap$ □haschild.human
Beast $\sqsubseteq$ animal $\sqcap \neg$ human
Woman $\sqsubseteq$ human $\sqcap$ female
man $\sqsubseteq$ human $\sqcap \neg$ woman
father $\sqsubseteq$ man $\sqcap \geq 1$ haschild.human
mother $\sqsubseteq$ woman $\sqcap \geq 1$ haschild.human
grandfather $\sqsubseteq$ father $\sqcap \exists$ haschild.parent
grandmother $\sqsubseteq$ mother $\sqcap \exists$haschild. Parent
motherwith5children $\sqsubseteq$ mother $\sqcap \geq$5haschild.
human $\sqcap \leq$5haschild. human
predators $\sqsubseteq$ beast $\sqcap \forall$ hasnutrition.meat

Ontology2

parent $\sqsubseteq$ humain $\sqcap \exists$a_enfant.humain
femme $\sqsubseteq$ humain $\sqcap$ feminine
homme $\sqsubseteq$ humain $\sqcap \neg$ woman
papa $\sqsubseteq$ homme $\sqcap \exists$a_enfant.humain
maman $\sqsubseteq$ femme $\sqcap \exists$a_enfant.humain
grandpere $\sqsubseteq$ papa $\sqcap \exists$a_enfant.parent
grandmere $\sqsubseteq$ maman $\sqcap \exists$a_enfant.parent
meredeSenfant $\sqsubseteq$ maman $\sqcap$ < 5 a_enfant.humain $\sqcap$ > 5 a_enfant.humain

We expect the system to obtain the mapping bellow (*note that for clarity purpose, edges were not labeled with roles*)

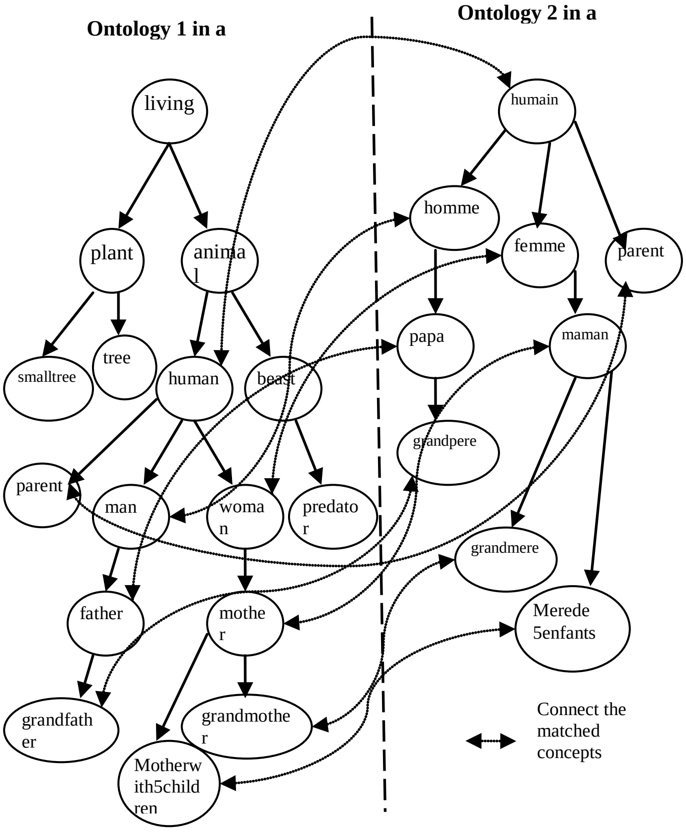


Fig.7 Ontologies represented in a graph with matched concepts

The system uses a threshold value to determine similarities between concepts. The comparison between concepts of ontology O_1 and concepts of ontology O_2 provides a tabular representation of the entire possible matching and their similarity values. Each table entry contains the two concepts being compared and their similarity score. Each concept has (if exist) only one similar concept in the second tree. A threshold value is set, which gives the minimum value at which, two concepts can be said to be semantically similar. Since the score of similarity between concepts is in the range of 0 to 1, the threshold is chosen as a positive value between 0 and 1.

The system scans the entire table to determine the table entry having the highest value. This value is compared with the threshold. If it is greater than the threshold, the corresponding concepts are taken to be the most similar concepts. For each table entry having one of the two concepts, they are removed from the table and therefore no other concept can be mapped to them. The process is iterated until no value of similarity obtained is above the threshold value.

Figures 8 and 9 below show the result for the two Ontologies given above. The first result evaluated without syntactic comparison and the second one with syntactic comparison. From the result it is clear that the syntactic comparison has introduced a discrepancy in the result which is increasing as comparison move downward the tree.

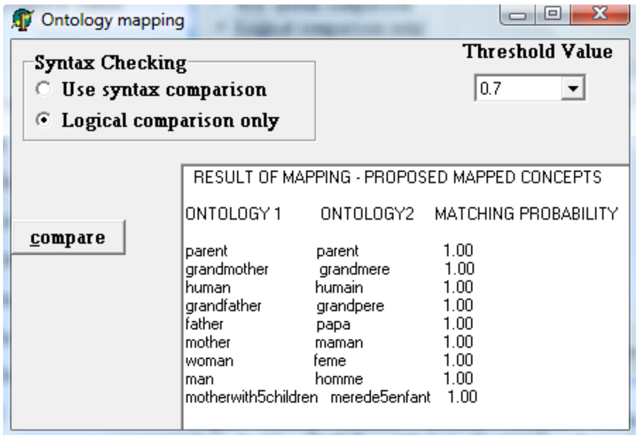


Fig.8 Output with similarity evaluated mainly on logical comparison(threshold = 0.7)

A comparison was performed on the same ontology while taken into account the syntactic similarities evaluation and the following result was obtained.

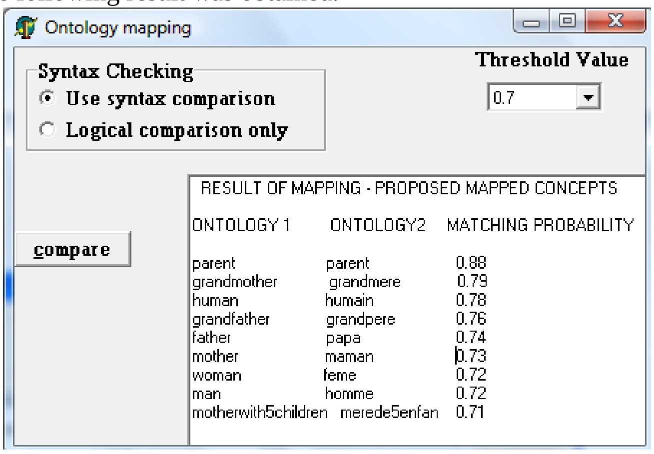


Fig.9 Output with similarity evaluated based on logical and syntactic comparison (threshold= 0.7)

6. Summary and Conclusion

This work proposes and evaluates a structural approach, based on SHIQ description logics language that facilitates interoperability between Ontologies. Textual Ontologies to be compared are translated into hierarchical trees of concepts and roles based on SHIQ description GCI. Similarities between concepts are evaluated with more emphasis on their logical definition and by referring to others related concepts.

The success of semantic interoperability largely depends on the possibility of system to operate less with the semantic similarities obtained from the syntactic analysis of information they are processing. It is hoped that, new intelligent application will deviate from syntatic to logical nature of semantic comparison. The framework evaluated in this work, shows that an analytical of semantic similarities between Ontologies based on their logical definition may be the expected solution to overcome the problems of interoperability encountered in Ontologies built in SHIQ. The proposed work is limited only on SHIQ language. SHIQ is just one of the multitude incompatible Ontology languages. Automatic processing of semantic information will not be fully achieved as long as we have several incompatible systems. It becomes then a necessity to look at performing a mapping across different plate form by finding techniques to reconcile existing ontology languages and by building theories that define the principle of mapping.

References

[1] Baader, F. and Nutt, W. Basic description logics. The Description Logic Handbook: Theory, Implementation, and Applications, 2003. pp 43–95. Cambridge University Press

[2] Baader, F., Horrocks Ian and Ulrike Sattler "Description logics as ontology languages for the semantic web." Mechanizing Mathematical Reasoning: Essays in Honor of Jörg Siekmann on the Occasion of His 60th Birthday, volume 2605 of Lecture Notes in Artificial Intelligence, 2005, Pp 228-248

[3] Baader, F., Horrocks Ian, Ulrike Sattler "Description Logics for the Semantic Web". KI - Künstliche Intelligenz, 2002 pp 57-59

[4] Ehrig Marc and Steffen Staab " QOM: Quick Ontology Mapping." International Semantic Web Conference, 2004 pp 683– 697

[5] Horrocks Ian "Reasoning with expressive description logics: Theory and practice" Proceeding of the 19th International Conference on Automated Deduction (CADE 2002), number 2392 in Lecture Notes in Artificial Intelligence, pp 1-15

[6] Horrocks Ian, Ulrike Sattler "Decidability of SHIQ with Complex Role Inclusion Axioms" journal of Artificial Intelligence, vol. 160(1-2) ,2004, pp79-104

[7] Horrocks Ian, Ulrike Sattler. "A description logic with transitive and inverse roles and role hierarchies." Journal of Logic and Computation, vol. 9(3),1999, pp385–410

[8] Horrocks Ian, Ulrike Sattler, Tobies, S. " Reasoning with individuals for the description logics SHIQ" Proceeding. of the 17th International Conference on Automated Deduction (CADE 2000), number 1831 in Lecture Notes in Artificial Intelligence, pp 482-496

[9] John Li "LOM: A Lexicon-based Ontology Mapping Tool". Proceeding of the Performance Metrics for Intelligent Systems (PerMIS.'04). Information Interpretation and Integration Conference.

[10]Madhavan, J., Bernstein, P. A., Domingos, P. and Halevy, A. "Representing and reasoning about mappings between domain models" Proceedings of the Eighteenth National Conference on Artificial Intelligence and Fourteenth Conference on Innovative Applications of Artificial Intelligence (AAAI 2002), pp 80–86, Edmonton, Alberta, Canada.. AAAI Press.

[11]Namyoun Choi, Il-Yeol Song, and Hyoi Han "A Survey on Ontology Mapping**"** SIGMOD, Vol. 35, No, 3, Sep. 2006, pp. 34-41

[12]Nardi, D., Brachman, J. An Introduction to Description Logics The Description Logic Handbook: Theory, Implementation, and Applications. Cambridge University Press, 2003, pp43–95

[13] Noy, N. F. and Musen M. A." Prompt: algorithm and tool for automated ontology merging and alignment" Proceeding of Seventeenth National Conference on Artificial intelligence (AAAI-2000)

[14]Noy, Natalya F "Semantic Integration: A Survey of Ontology Based Approaches." Stanford Medical Informatics, 2004. Stanford University.

[15]Uschold Mike, Michael Gruninger "ontologies: principles, methods and applications" Knowledge Engineering Review, vol11(2) 1996, pp93-155

[16]Valente, A., Russ, T., MacGrecor, R. and Swartout, W. "Building and (re)using an ontology for air campaign planning" IEEE Intelligent Systems, 1999, pp 27

[17]Yannis Kalfoglou1 and Schorlemmer, M. "Ontology mapping: the state of the art." The Knowledge Engineering Review, Vol. 18(1), 2003pp1–31.Cambridge University Press

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Parallel Genetic Algorithm System

Nagaraju Sangepu
Assistant Professor
nag_sangepu@yahoo.com

K.Vikram
CSE dept, KITS
Warangal, India
vikram.kalwala@gamil.com

ABSTRACT – Genetic Algorithm (GA) is a popular technique to find the optimum of transformation, because of its simple implementation procedure. In image processing GAs are used as a parameter-search-for procedure, this processing requires very high performance of the computer. Recently, parallel processing used to reduce the time by distributing the appropriate amount of work to each computer in the clustering system. The processing time reduces with the number of dedicated computers. Parallel implementations of systems can be grouped into 3 categories: 1) Parallel Hardware architectures designed specially for parallel processing. 2) Supporting Software implementations on machines with hardware support for parallel processing and 3) parallel processing algorithms implemented entirely in software on general-purpose hardware. It includes the clustering architecture consisting of homogeneous collection of general purpose computer systems connected via networks, also termed a clustered computing environment. . The queue length is optimally adjusted using GA so that queue length is minimized during data transfer in order to keep the bandwidth at a stable condition. Graph is also drawn to show the difference of bandwidth. Implementation includes all those activities that take place to convert from the old system to the new. The new system may be totally new, replacing an existing system or it may be major modification to the system currently put into use. This application implemented with simulation model of computer network, constructed along with the router. The options are given to invoke the FCFS and Genetic Algorithm. The path between source and destination were drawn and the result of both algorithms is discussed.

1. INTRODUCTION

GA mimics all the process based on the concept of natural evolution to find the optimized solution to the given problem residing in the search space. The GA pool contains a number of individuals called chromosomes. Each chromosome encoded from the parameters holds the potential solution. According the evolutionary theories, the chromosomes which only have a good fitness are likely to survive and to generate the off springs and pass its strength to them by the genetic operator. The fitness of chromosome is the way that is linked to the predefined problem or objective function. Fig 1 shows the possible stages of evolution.

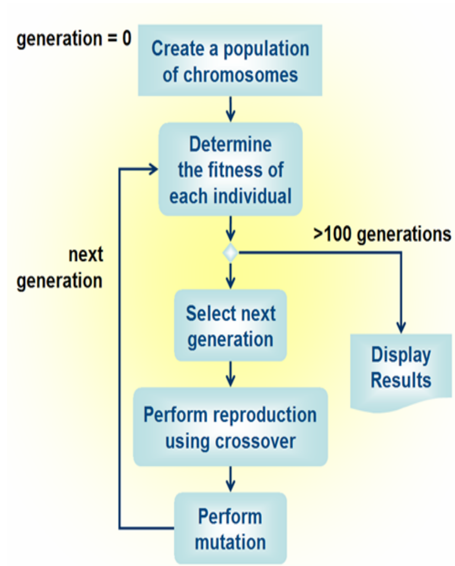


Figure 1. Genetic Algorithm Steps

GA cycle can be decomposed into five steps described as follows:

1) Randomly initialize the population in the pool. With more population, the coverage in search space is good but traded off by the calculation time in each generation. In the simplest way, the real – value parameter is binary coded to give a bit string. The bit strings for several parameters are concatenated to form a single string or chromosome. In accord with the biology, each bit corresponds to gene.

2) Evaluate the chromosomes by objective function. After the evolution, all the chromosomes are ranked for the fitness values in the descending or ascending order depending on the purpose of objective function.

3) Select the parents from the chromosomes with the based chances. The higher-fitness chromosome is prone to survive.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

4) Generate the offspring using genetic operators consisting of crossover and mutation. Crossover is a recombination operator that swaps the parts of two parents, two random decisions are made prior to this operation, whether to do it or not and where the crossover point is. Mutation gives a good chance to explore the uncovered search space. It mutates, or complements some genes in the chromosome of the offspring, so that the new parameter value takes place.

5) Entirely replace the elder generation in the pool with the newer one and return to step 2. In some case, the few best elders may be kept away from replacement. This is known as elitist strategy. The criteria for stopping the revaluation loops are met when a (the loop number is over some predefined point or d) the steady last for predetermined times.

It is important to realize that GA are stochastic meaning that there is randomness involved ; mainly in the initial generation of a random population, random choice of parents, random choice of which genes to inherit from which parents, and random choice of which genes to mute. Some times how however tournament selection is used as a parent or survival selection strategy. This ensures that there is bias towards replacing less fit solutions in the parent population by fitter solution from the new generation.

2. PARALLEL GENETIC ALGORITHM

Two approaches to parallel genetic algorithms have been considered so far.

Standard parallel approach: In this approach, the evaluation and the reproduction are done in parallel. However, the selection is still done sequentially, because would require a fully connected graph of individuals in the population may be mated.

Decomposition approach: This approach consists in dividing the population into equal size sub-populations. Each processor runs the genetic algorithm on its own sub-population, periodically selecting good individuals to send its neighbors and periodically receiving copies of its neighbors' good individuals to replace bad ones in its own sub-population.

The processor neighborhood, the frequency of exchange and the number of individuals exchanged are adjustable parameters. The standard parallel model is not flexible in the sense that the communication overhead grows as the square of population size. Therefore, this approach is not adapted to distributed memory architectures, where the cost of communications has a great impact on the performance of parallel programs. In the decomposition model, the inherent parallelism is not fully exploited as treatment of sub-populations may be further decomposed. This approach should be considered only when the number of available processors is less than the required size of the population. Considering massively parallel architectures with numerous processors, we chose a fine –grained model, where the population is mapped on a connected processor graph like a grid, one individual per processor. We have bisection between the individual set and the processor set. The selection is done locally in a neighborhood of each individual. Another version to this approach has been already proposed in, where at each generation a hill-climbing algorithm is executed for each individual in the population. The choice of the neighborhood is the adjustable parameter. To avoid overhead and complexity of routing algorithms in parallel distributed machines, a good choice may be restrict neighborhood to only directly connected individuals.

The parallel genetic algorithm proposed is: Genetic in parallel a population of random individuals.

While *generation_number <max_number_of_generations*
Do

1. *Evaluation*- Evaluate in parallel each individual.
2. *Selection* – Receive in parallel the individuals coming from the neighbors.
3. *Reproduction*- Each individuals reproduces in parallel with the individuals previously received.
4. *Replacement*- Do in parallel a selection of best local off springs.

We use the speed-up ratio as a metric for the performance of the parallel genetic algorithm. The speed up ratio S is defined as $S=T_s/T_p$ where T_s is the execution time on a single processor and T_p corresponds to execution time for a P processors implementation.

3. RELATED WORK

There are mainly two types of scheduling namely the system level scheduling and the application level scheduling. The scheduling system will analyze the load situation of every node and select one node to run the job. The scheduling policy is to optimize the total performance of the whole system. If the system is heavily loaded, the scheduling system has to realize the load balancing and increase the throughput

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

and resource utilization under restricted conditions. This kind of scheduling is known as the system level scheduling.

If multiple jobs arrive within a unit scheduling time slot, the scheduling system shall allocate an appropriate number of jobs to every node in order to finish these jobs under a defined objective. Obviously, the objective is usually the minimal average execution time. This scheduling policy is application-oriented so we call it application-level scheduling.

We will first select a certain number of inputs, say, x_1, x_2, ... , x_n belonging to the input space X. In the GA terminology, each input is called an *organism* or *chromosome*. The set of chromosomes is designated as a *colony* or *population*. Computation is done over *epochs*. In each epoch the colony will grow and evolve according to specific rules reminiscent of biological evolution.

To each chromosome x_i, we assign a fitness value which is nothing but $f(x_i)$. Stronger individual that is those chromosomes with fitness values closer to the colony optimal will have greater chance to survive across epochs and to reproduce than weaker individuals which will tend to perish. In other words, the Algorithm will tend to keep inputs that are close to the optimal in the set of inputs being considered (the colony) and discard those that under-perform the rest.

The crucial step in the algorithm is reproduction or breeding that occurs once per epoch. The content of the two chromosomes participating in reproduction are literally merged together to form a new chromosome that we call a child. This heuristic allows us to possibly combine the best of both individuals to yield a better one (evolution).

During each epoch, a given fraction of the organisms is allowed to mutate. This provides a degree of randomness which allows us to span the whole input space by generating individuals with partly random genes.

Each epoch ends with the deaths of inapt organisms. We eliminate inputs exhibiting bad performance compared to the overall group. This is based on the assumption that they're less inclined to give birth to strong individuals since they have poor quality genes and that therefore we can safely disregard them (selection).

The algorithm: Now that we've outlined the basic principles, let's examine in further detail how this whole process is accomplished and how the algorithm works in practice. Let's take the example of optimizing a function f over a space X

Every input x in X is an integer vector x=(x_1, x_2, ... , x_n). For the sake of simplicity, assume $0 <= x_i <= k$ for i=1...n. In order to implement our genetic algorithm for optimizing *f*, we first need to encode each input into a chromosome.

We can do it by having *log(k)* bits per component and directly encoding the value x_i (figure 1). Each bit will be termed *gene*. Of course, we may choose any other encoding based on our requirements and the problem at hand.

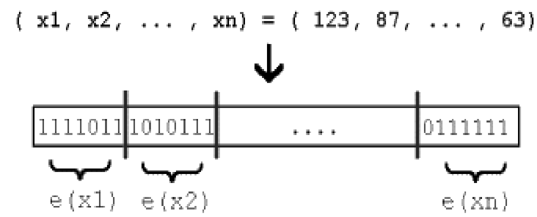


Figure 2. Encoding the Chromosomes

At epoch 0, we generate (possibly randomly) an initial set of inputs in X. Then at each epoch i, we perform fitness evaluation, reproduction, mutation and selection. The algorithm stops when a specified criterion providing an estimate of convergence is reached.

Reproduction: At each epoch, we choose a set of chromosomes belonging to the population that will mate. We choose to call such individuals females. Each female chooses a random set of potential partners and mates with the fittest of the group (this is another way of achieving selection). Once two organisms have been chosen for crossover, we merge their genetic information in order to create a new organism. The split position is determined randomly.

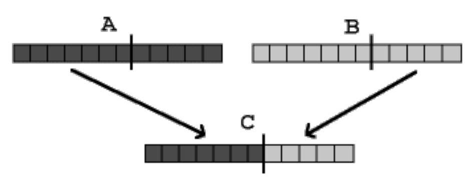


Figure 3. CrossOver Operation

Mutation: A new organism is created by randomly modifying some of its genes. This can be done right after reproduction on the newly created child or as a separate process.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

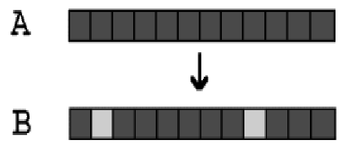


Figure 4. Mutaion Operation

Death: Worst performers among the colony are given a high probability of dying at the end of each epoch. We may also consider eliminating old chromosomes. The highest performer is immune from death from old age.

Why do Genetic Algorithms Work?

Similarities among the strings with high fitness value suggest a relationship between those similarities and good solutions.

A *schema* is a similarity template describing a subset of strings with similarities at certain string positions. Crossover leaves a schema unaffected if it doesn't cut the schema. Mutation leaves a schema unaffected with high probability (since mutation has a low probability). Highly-fit, short schemas (called building blocks) are propagated from generation to generation with high probability.

Competing schemata are replicated exponentially according to their fitness value. Good schemata rapidly dominate bad ones. The effectiveness of the search depends on the population size and the number of generations. The larger the population, the more likely that our initial population is representative of the search space, and the more likely that a probabilistic survival of the fittest mechanism produces the expected outcomes. Each successive generation should improve the fitness of the result, so longer runs usually produce better solutions. Genetic Algorithm application level scheduling algorithm generates the initial population, evaluates each individual's fitness, and performs genetic operations on the individuals with high fitness such copying, crossover and mutation, to generate a new population. The genetic process continues with the new population until a nearly optimal jobs assignment strategy is obtained. Finally, the jobs are assigned to each node based on the strategy. The connection to a resource is limited and a limited service is provided to the jobs. The scheduling policies used are the greedy algorithm which assigns the resources as and when it is found. Suppose that there are three data servers {S1, S2, S3}, each having two available connections. Let S1 have resources {r1, r2, r3, r4} and both S2 and S3 have resources {r1,r2,r5,r6}. Suppose the scheduler has four tasks each processing one of the resources. Each task with no contention, run for one hour. A greedy scheduler could allocate the two connections of S1 for running the resources r1 and r2. The running time is two hours as the other tasks cannot be run. The parameters to be considered in job scheduling are the following.

- Total execution time is the time between the beginning of execution of the first job of a series and completion of the last job.
- Average turnaround time is the average, for each job from when the job arrives to when the job finished.

Parallel Genetic Algorithms PGA has the same advantage as a serial Genetic algorithm, consisting in using representation of the problem parameters, robustness, easy customization for a new problem and multiple solution capabilities. PGA is usually faster, less prone to finding only sub-optimal solutions, and able of cooperating with other search techniques in parallel. PGA can be divided into global, fine grained, coarse grained and hybrid models.

The advantages of using PGA as stated in are

- Parallel search from multiple points in a space
- Works on a coding of the problem.
- Independent of the problem.
- Can yield alternate solutions to the problem.
- Better search even if no parallel hardware is used.
- Higher efficiency and efficacy than sequential Genetic Algorithms.
- Easy cooperation with other search procedures.

The global single population master slave Genetic algorithm tells the master stores the population, executes the Genetic operators, and distributes individuals to the slaves. The slave evaluates the fitness of the individual and reports the fitness value to the master.

4. PROPOSED SYSTEM

Here the problem statement is to reduce the processing overhead during scheduling. In our implementation, we apply the proposed work to data transfer between computers of two networks. Generally, during data transfer between pc of two

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

different networks, a router will be present in between the networks and it will take care of the scheduling of data packets between the source and destination computers. In the router there will be a number of ports P_n and each port P_i will take care of one data transfer. In each port, there will be a queue Q_i for data packets and this is where scheduling is applied. There are various scheduling algorithms possible to schedule the packets in each port of the router. The objective of each router is to reduce the congestion *C* of data transfer. Here we compare the proposed method with FCFS (first-come-first-serve) scheduling. We show the difference in terms of bandwidth at the router. Bandwidth will be kept in a stable condition and hence possibility of congestion and deadlock are greatly reduced. The queue length *ql* is optimally adjusted using GA so that queue length *ql* is minimized during data transfer in order to keep the bandwidth at a stable condition. Graph is also drawn to show the difference of bandwidth BW. Implementation includes all those activities that take place to convert from the old system to the new. The new system may be totally new, replacing an existing system or it may be major modification to the system currently put into use. This application implemented with simulation model of computer network, constructed along with the router. The options are given to invoke the FCFS and Genetic Algorithm. The path between source and destination were drawn and the result of both algorithms is discussed.

5. IMPLEMENATION

It is necessary to select a certain number of inputs space X. Let *BW* be the total band width available, *Ql* is the queue length, there are *n* queues belonging to input space X. In the GA terminology, each input is called an organism or chromosome. The set of chromosomes is designated as a colony or population.

Every input *x* in X is an integer vector x=(x_1, x_2, $x_{3...}$ x_n).

Genetic in parallel a population of random individuals

While SUM(all queue length) <= total bandwidth *BW*

For values of X 0<=xi<=k , *for i=1...n*;

As the band width is constant and the queue length is variable, kept less than bandwidth (Q <= BW)

At epoch 0, we generate (possibly randomly) an initial set of inputs in X.

Then at each epoch *i*, call *genetic_alg(),*

Repeat above steps for all epochs

The algorithm *genetic_alg()* performs the following operations of GA

a. fitness evaluation,

b. reproduction,

c. mutation and

d. selection.

The algorithm stops when a specified criterion providing an estimate of convergence is reached.

Fitness Function evaluation can be on either maximization of Bandwidth *(BW)* or minimization of Congestion *C*.

6. RESULT

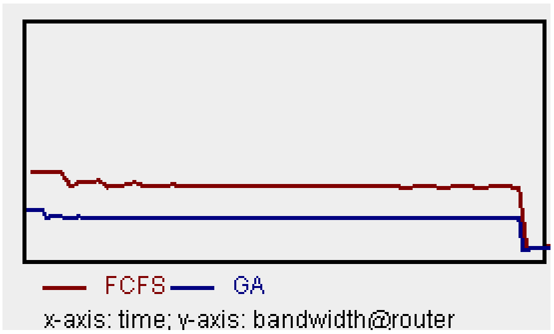


Figure 5. Results FCFS Vs GA

The main results are the following:
- The algorithm has higher speed
- It is easy to program
- It is simple to implement on massively parallel distributed memory architectures.

Genetic algorithms are versatile optimization tools suitable for solving multi-disciplinary optimization problems in aerodynamics where the design parameters may exhibit multimodal or non-smooth variations. The ability of genetic algorithms (GAs) to handle non- smooth topology and overcome local critical points in its search for a global optimum makes the method ideal for use in aerospace design optimization.

7. CONCLUSION

Here the problem statement is to reduce the processing overhead during scheduling. In our implementation, we apply the proposed work to data transfer between computers of two networks. Generally, during data transfer between pc of two different networks, a router will be present in between the networks and it will take care of the

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

scheduling of data packets between the source and destination computers. In the router there will be a number of ports and each port will take care of one data transfer. In each port, there will be a queue for data packets and this is where scheduling is applied. There are various scheduling algorithms possible to schedule the packets in each port of the router. The objective of each router is to reduce the congestion of data transfer. Here we compare the proposed method with FCFS (first-come-first-serve) scheduling. We show the difference in terms of bandwidth at the router. Bandwidth will be kept in a stable condition and hence possibility of congestion and deadlock are greatly reduced. The queue length is optimally adjusted using GA so that queue length is minimized during data transfer in order to keep the bandwidth at a stable condition. Graph is also drawn to show the difference of bandwidth.

8. References

[1] Yang, Hongquiang, Joshua, Adaptive grid job scheduling with genetic algorithm, *Elsevier Future Generation Computer Systems 2005.*

[2] D.E Goldberg, "*Genetic Algorithm in Search, Optimization and Machine Learning* ", Prentice Hall of India, 2004.

[3] *Genetic Algorithms: A Survey* by M. Srinivas and Lalit M. Patnaik, IEEE Computer, v. 27, 6, June 1994, pp. 17-27.

[4] Alessandro, Keith, "Multiple job Scheduling in a Connection limited Data Parallel System", *IEEE Transactions on parallel and Distributed Systems. vol 17, no 2 Feb 2006*

[5] P.Visalakshi, T.Hamsapriya, S.N.Sivanandam, "Parallel Genetic Algorithms-State of the art Survey" JCS vol. 1 No.6 May-June 2006.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Framework for vulnerability reduction in real time intrusion detection and prevention systems using SOM based IDS with Netfilter-Iptables

Abhinav Kumar
X-Scholar, CSE Department
Jaypee Institute of Information Technology, Deemed University
Noida, India
abhinavjiit@yahoo.co.in

Kunal Chadha
X-Scholar, CSE Department
Jaypee Institute of Information Technology, Deemed University
Noida, India
id.kunal@gmail.com

Dr. Krishna Asawa
Associate Prof., CSE/IT Department
Jaypee Institute of Information Technology, Deemed University
Noida, India
krishna.asawa@jiit.ac.in

*Abstract***— Intrusion detection systems and Intrusion Prevention system are few of the possible ways for handling various types of attacks or intrusions. But the credibility of such systems itself are at stake. None of the existing systems can assure you, your safety. In this paper we propose integration of SOM based intrusion detection system with an intrusion prevention system in the Linux platform for preventing intrusions. We propose a framework for reducing the real time security risks by using Self-organizing maps for intrusion detection accompanied by packet filtering through Netfilter-Iptable to handle the malicious data Packets.**

Keywords-Intrusion Detection System, SOM.

I. Introduction

In today's world every computer is vulnerable, nothing is secure, but the quest of mankind for that ideal security is still going on. Internet and other ways of communication over network are proving to be boon as well as bane. Boon, when it is providing new dimensions to the business and bane with its harmful effects of intrusions into various networks. Every now & then we witness various types of attacks and keep banging our heads in solving them. As soon as one computer is connected to another computer there is an addition of the possibility that someone using the other computer can access our computer's information, eventually leading to intrusions. Some recent surveys show that cyber attacks targeted to the networks are no longer an unlikely incident that only occurs to few exposed networks of organizations in the limelight. In the struggle to both maintain and implement any given IT security policy, professional IT security management is no longer able to ignore these issues, as attacks are more frequent and devastating; the commercial success is directly related to the safe and reliable operation of their networks [4].

Intrusion is an action to attack the integrity, confidentiality and availability of the system resources [3]. Intrusion detection systems were developed for this cause so that they can detect the malicious data packets traveling on the network in real time. But it has its own limitations such as it can't do the session based detection which uses multiple packets [2]. In a network based IDS, packets are examined both according to header and payload searching for attack signatures, stored in the IDS Attack signature database, which is the vital part of any IDS software [4] but it becomes inefficient when we talk about blocking those attacks and hence can easily enter into a system. Each of such system is passive in reporting such intrusions and hence do not provide real time security.

For handling such situations we propose a real time system that consists of an intrusion detection system based on Self organizing maps, for tracing down the malicious packets along with handling those packets through an intrusion prevention system in the Linux environment. Self-organizing maps is an unsupervised way of learning and has the ability to express topological relationships [22]. The hypothesis is that typical connection characteristics will be emphasized – densely populated regions of the map – whereas atypical activities will appear in sparse regions of the topology [22]. Selection of SOM for intrusion detection is also guided by its robustness with regard to the choice of the number of classes to divide the data into, and is also resistant to the presence of outliers in the training data, which is a desirable property: in the real-world situations, the training data could already contain attacks or anomalies and the algorithm must be capable of learning regular patterns out of a "dirty" training set [25]. Detection will be followed by prevention by using Netfilter-Iptables available in Linux environment [3]. Our system blocks the malicious data packet as soon as they are detected, without any external help, in real time.

This paper is organized in various sections in which we discuss the existing intrusion detection system as well as intrusion prevention systems. This is followed by description of framework which consists of training of SOM, usage of netfilter-iptables for packet filtering.

.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

II. EXISTING INTRUSION DETECTION SYSTEM

Scientist and researchers had been continuously working for quite a few years for the development of a perfect intrusion detection system (IDS) that can't be bluffed. Its main job is to monitor, analyze, detect and respond to the intrusions to the information systems [5]. Intrusion detection systems can be broadly categorized into signature based and anomaly detection systems. It may be passive. Signature based IDS look for attack signatures, specific patterns of network traffic or activity in log files that indicate suspicious behavior. Signature-based methods rely on a specific signature of an intrusion which when found triggers an alarm [6, 7]. Now coming on to its sub categories-if an IDS searches for suspicious attack signatures on the traffic flowing on the network then it is named as Network intrusion detection systems (NIDS) and when the same is done by looking at log file of hosts, it is termed as Host intrusion detection systems (HIPS) [4]. HIDS is mostly deployed in e-commerce environments for securing the sensitive data. But it serves the purpose only at the host level. NIDS performs the search for attack signatures at the packet level and as soon as a match is found, an alarm gets raised. The anomaly detection IDS uses statistical techniques to detect penetrations and attacks that begins with the establishment of base-line statistical behavior that what is the normal behavior for this system. After that it captures new statistical data and measure, for finding the deviation from the base line. Once a threshold is exceeded, an alarm is generated [4].

All the above-mentioned IDS's suffer from few serious limitations. As the attack-trails is increased, it became difficult for network IDS or host IDS to detect the attacks with a limited capability [9]. Some of them are 1) High misinformation rate-is a bulky log and real-time prevention problems that has not yet been solved efficiently [3]. An alarm gets raised even if there was no attack (false positive) and no alarm even if there is as an attack (false negative). Hence there is need for a more exact and effective access control policy [8]. Hence in anomaly detection methods, the base line needs to be adjusted dynamically. 2) Once an IDS gets attacked then it allows the attacker to move freely on the network [8] .3) There is no way by which an IDS can block an attacker, it remains confined only to its primary job of detection.

III. INTRUSION PREVENTION SYSTEM

Intrusion prevention system (IPS), also known as Network Defense System (NDS), is a system in which firewall is tightly coupled with IDS and it can react to the changes of the network environment [8]. It can be either in the form of software or hardware providing help in blocking of illegal external attack, preventing the loss, destroy and change of internal information from illegitimate users through Internet, and helping internal information to be provided to the outside safely. It is an active protection process to prohibit from incoming of illegal traffics and permit only the authorized traffics [17]. IPS is located in the rear section of router generally and keeps a check on the forwarded packets to the router by analyzing and comparing with filter-rules [16]. In order to have proper security the IPS should fulfill the criterions like- it must be a part of communication link and supported by dedicated hardware, it should actively detect the intrusions in real time and should block those intrusions instantaneously.

IV. PROPOSED FRAMEWORK

The proposed framework for efficient intrusion detection-protection system is an integration of SOM based intrusion detection system working in coherence with netfilter-iptable based firewall. Self Organizing maps being an unsupervised way of learning are one of the best choices for intrusion detection because it clearly identifies the "odd" phenomenon even in vast amount of observations, which is its core property. Apart from this, it does not require a priori knowledge inputs.

The DARPA 1998 Intrusion Detection Evaluation data set consists of about 5 million connections of labeled training data and 2 million connections of test data [23]. This data consists of the values of all 41 features of a data packet along with its labeling into categories of normal, smurf, Neptune etc. These 41 features consist of Basic TCP features, Content features, Time-based traffic features; and Host-based traffic features [24]. Since the proposed work is data driven unsupervised from of learning hence out of those 41 features only 6 having basic TCP information are required, namely-duration of connection, protocol type (tcp/udp), service(HTTP etc.), destination bytes, source bytes and the value of flag. Hence SOM based IDS will have 6 inputs and classifies packets into three clusters-normal, smurf and Neptune, the latter two being the attacks. Once this network gets trained with this data, it is ready for detecting the malicious packets.

- *Why SOM for intrusion detection?*

Intrusions done by an unknown program leads to disasters because of their unknown behavior & characteristics. Although we can find out its characteristics but they remain a mystery for us. So we need to classify it into the normal and the abnormal states [11]. Now the problem gets reduced to defining normal and the abnormal states.

The architecture of Self organizing maps was developed by Teuvo Kohonen at the University of Helsinki. Self organizing maps are provided only with a series of input patterns and it learns for itself how to group these together so that similar patterns produce similar outputs. It consists of a single layer of cells, all of which are connected to a series of inputs. The inputs themselves are not cells and do no processing - they are just the points where the input signal enters the system [14] as shown in Figure 1.

.

 http://sites.google.com/site/ijcsis/
ISSN 1947-5500

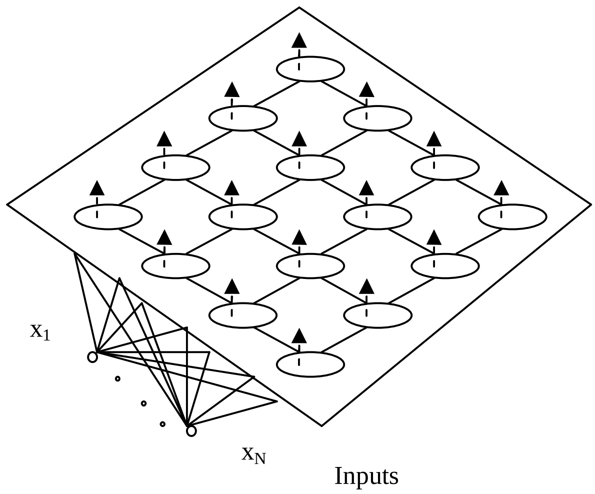


Figure 1 (Self Organizing Map)

This network involves unsupervised learning and hence it itself finds, what it needs to learn, without any external help. In the area of intrusion detection systems, the use of unsupervised learning algorithms supports the detection of anomalies [10, 12]. Moreover a learning algorithm can be tuned totally to the specific network it operates into, which is also an important feature to reduce the number of false positives and optimize the detection rate [12].

- *Training the SOM*

The training of self organizing map involves sampling, similarity matching and updating apart from the basic initialization of weights to very small values of the range 0 to 0.01 [13]. The learning process of SOM is as follows:

1) During initialization, the only restriction is that $\mathbf{w}_\mathbf{j}$ (0) be different for $j=1,2,\ldots l$, where l is the number of neurons in the lattice.

2) It is followed by sampling where a sample vector x (representing activation pattern) is drawn from the input space with certain probability and presented to the lattice. In the proposed work, out of previously mentioned 41 features, the 6 basic TCP information are presented to the network.

3) In similarity matching every node is examined to calculate which one's weights are most like the input vector. The winning node is commonly known as the Best Matching Unit (BMU)/neuron. BMU is calculated by iterating through all the nodes and calculating the Euclidean distance between each node's weight vector and the current input vector. Hence the BMU $i(\mathrm{x})$ at time step n by using the minimum –distance Euclidean criterion is:

$$\boldsymbol{i}(\mathrm{x})=\arg\min_{\mathrm{j}} \| \mathbf{x}(n) - \mathbf{w}_{\mathrm{j}}\|, \boldsymbol{j}=1,2,\ldots, l$$

---------------- Formula 1

The node with a weight vector closest to the input vector is tagged as the BMU. As the learning proceeds and new input vectors are given to the lattice, the learning rate gradually decreases to zero according to the specified learning rate function type [15]. Along with the learning rate, the neighborhood radius decreases as well.

4) In the updating phase the synaptic weight vectors of all the neurons is updated by using the formula

$$\mathbf{w}_{\mathrm{j}}(n + 1) = \mathbf{w}_{\mathrm{j}}(n) + \mathbf{n}(n)\, \boldsymbol{h}_{j,i(\mathrm{x})}(n)\, (\mathbf{x}(n) - \mathbf{w}_{\mathrm{j}}(n))$$

--------------- Formula 2

where $\mathbf{n}(n)$ is the learning-rate parameter, which has been set to 0.1 and $\boldsymbol{h}_{j,i(\mathrm{x})}(n)$ is the neighborhood function centered around the winning neuron $\boldsymbol{i}(\mathrm{x})$; both $\mathbf{n}(n)$ and $\boldsymbol{h}_{j,i(\mathrm{x})}(n)$ are varied dynamically during learning for best results [14].

5) We continue with step 2 until no noticeable changes in the feature map are observed or for given number of iterations (generally is fixed, in our case it is 50000).

After training, SOM becomes ready to categorize the packets in three different categories, namely-smurf, Neptune and Normal. After this phase the work of Intrusion prevention system starts. The efficiency of IPS gets decreased because of certain limitations in its basis principles. IPS performs packet filtering based on predefined rules, what if there is a novel attack? IPS has passive characteristics such that it can prevent only the predefined rules and filter some kinds of packets [18]. Apart from this, it is also not able to detect an attack carried out from the internal network of an organization. We propose to use Netfilter-Iptables for overcoming many such drawbacks of intrusion prevention systems.

- *Netfilter-Iptable*

Netfilter is a set of hooks inside the Linux kernel [18]. Netfilter is a framework that enables packet filtering, network address [and port] translation and other packet mangling. It performs packet filtering based on rules saved in packet filtering tables in kernel space. The rules are grouped in chains, according to the types of packets they deal with. Rules dealing with incoming packets are added to the INPUT chain, rules dealing with outgoing packets are added to the OUTPUT chain and rules dealing with packets being forwarded are added to the FORWARD chain [20]. Other than these three chains there are other chains also like prerouting & postrouting and user defined chains. As soon as a packet comes to a chain, its next action is decided on that chain.

When a packet perfectly matches with a rule, action performed is ACCEPT and packet is allowed to go wherever it is destined to(-j ACCEPT), DROP-packet will be blocked and no further processing will be done on it (-j DROP), REJECT(similar to drop) but doesn’t leave dead sockets & sends back error message (-j REJECT) as shown in Figure 2 [21]. There are few more actions that can be performed on the packets.

.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

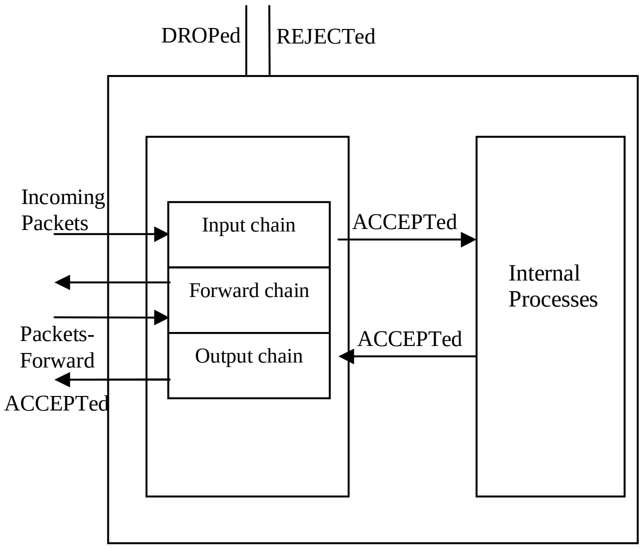


Figure-2 (Netfilter system) [21]

The iptables tool inserts and deletes rules from the kernel's packet filtering table [21]. 'Iptables' is not a packet filtering tool itself. It is just a command tool of the Netfilter imported in the kernel, and we should use this tool to make rules to reflect current intrusion aspects [3]. Few of its commands are: (-N) Creation of new chain, (-L) List the rules in a chain, (-F) Flush the rules out of a chain, (-A) Append a new rule to a chain,(-I) Insert a new rule at some position in a chain, (-X) delete an empty chain, (-D) delete a rule at some position in a chain, or the first that matches etc. For example for deleting the rule1

 # iptables -D INPUT 1

For blocking an IP address 192.168.1.1

iptables -A INPUT -f -d 192.168.1.1 -j DROP

Now as soon as the SOM based IDS find an attack it generates an alert. Along with generating an alarm it also passes the information, the port number and IP address of that malicious packet to the netfilter-iptable firewall. Then the IPS (firewall) decides how to deal with that packet according to the rules of the kernel's packet filtering table. The decision regarding dropping, accepting or rejecting the incoming packets is taken at this juncture after matching the packets with the predefined rules present in various chains (input, output, forward). And in cases of indecision or if any rule is not present in packet filtering table, it updates the table by inserting additional rules into it. This property of Netfilter-Iptable overcomes its limitation of handling only such packets for which predefined rules are available. This updation in the rules table is carried out by using libiptc (libiptc is a library to set the packet filtering function in the Netfilter framework) [3] and can be in the form of blocking that particular IP address or blocking only that particular port as shown in Figure 3.

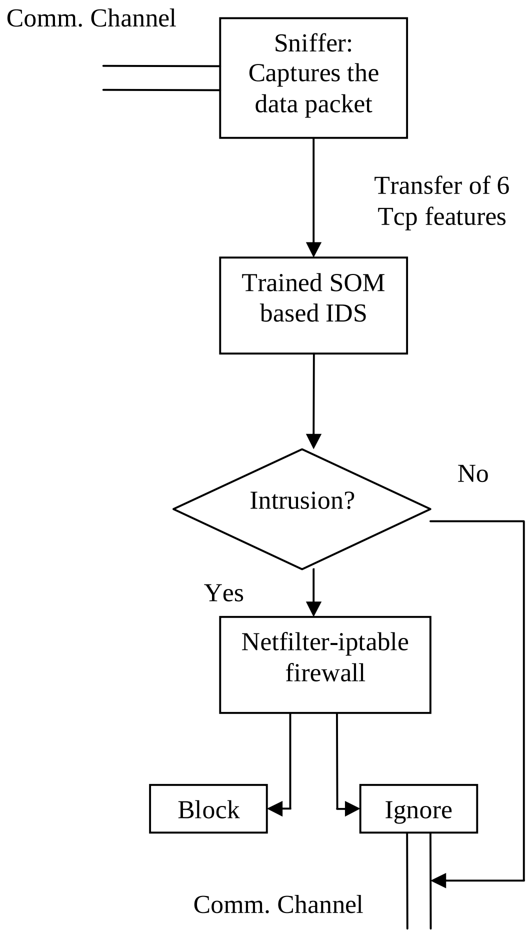


Figure-3 (Integrated Framework)

V. CONCLUSION AND FUTURE WORK

In this research we have investigated few of the intrusion detection and prevention systems and critically analyzed them. We have explored the role of self organizing maps, an artificial intelligence technique for increasing the efficiency of intrusion detection systems. We also presented an extensive study of Netfilter-Iptable for overcoming few of the limitations of existing intrusion prevention systems. Along with this we finally proposed an integrated version of SOM based IDS with netfilter-iptable firewall that do not require any external help in form of administrator for handling the malicious data packets. During the research we focused only on three classes-normal, smurf and Neptune. More practical

.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

IDSs should have several attack types; therefore, it is possible, as a future development to the present study, to include more attack scenarios in the dataset. We have taken only 6 basic tcp information of a packet for training our network for intrusion detection. Hence in future further improvements can be done by including more parameters of a data packet.

REFERENCES

[1] Kulesh Shanmugasundaram, Nasir Memon, Anubhav Savant, and Herve Bronnimann. ForNet: A Distributed Forensics Network. V. Gorodetsky et al. (Eds.): MMM-ACNS 2003, LNCS 2776, pp. 1–16, 2003.c _ Springer-Verlag Berlin Heidelberg 2003

[2] Bace, R.G.: Intrusion Detection. Macmillan Technical Pub (2000)

[3] Min Wook Kil, Seung Kyeom Kim, Geuk Lee and Youngmi Kwon. A Development of Intrusion Detection and Protection System Using Netfilter Framework. D. ´Slezak et al. (Eds.): RSFDGrC 2005, LNAI 3642, pp. 520–529, 2005.c_Springer-Verlag Berlin Heidelberg 2005

[4] Andreas Fuchsberger. Intrusion Detection Systems and Intrusion Prevention Systems. 1363-4127 Published by Elsevier Ltd.doi:10.1016 / j.istr.2005.08.00, 2005

[5] Jeong, B.H., Kim, J.N., Sohn, S.W.: Current Status and Expectation of Techniques for Intrusion Protection System. http://kidbs.it.nd.or.kr/WZIN/ jugidong/1098/109801.htm

[6] Ilgun, K., Kemmerer, R.A., and Porras, P.A.: State transition analysis: A rule based intrusion detection approach. IEEE Transactions on Software Engineering (March 1995)

[7] Kumar, S. and Spa.ord, E.H.: An application of pattern matching in intrusion detection. Purdue University Technical Report CSD-TR-94-013 (1994)

[8] Xinyou Zhang, Chengzhong Li and Wenbin Zheng. Intrusion prevention system design. 0-7695-2216-5/04. IEEE(2004)

[9] Shim, D.C.: A trend of Intrusion Detection System. KISDI IT FOCUS 4. Korea Information Strategy Development Institute (2001) 61-65

[10] U. Labib and V. R. Vemuri. Nsom: A tool to detect denial of service attacks using self-organizing maps.

[11] Sahin Albayrak, Achim Muller, Christian Scheel and Dragan Milosevic. Combining Self-Organizing Map Algorithms for Robust and Scalable Intrusion Detection. Proceedings of the 2005 International Conference on Computational Intelligence for Modelling, Control and Automation, and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'05), 2005.

[12] Stefano Zanero and Sergio M. Savaresi. Unsupervised learning techniques for an intrusion detection system. *SAC'04* March 14-17 2004, Nicosia, Cyprus Copyright 2004 ACM 1581138121/ 03/04.

[13] http://richardbowles.tripod.com/neural/kohonen/kohonen.htm

[14] Haykin, Simon: Neural networks- a comprehensive foundation. Pearson Education (4^{th} Indian reprint, 2003)

[15] Liberios Vokorokos, Anton Balaz and Martin Chovanec. Intrusion detection system using self organizing map. Acta Electrotechnica et Informatica No. 1, Vol. 6, 2006

[16] Min Wook Kil, Si Jung Kim, Youngmi Kwon and Geuk Lee. Network Intrusion Protection System Using Rule-Based DB and RBAC Policy. IFIP International Federation for Information Processing, NPC 2004, LNCS 3222, pp. 670-675, 2004.

[17] http://www.terms.co.kr, Dictionary of Computer Terms.

[18] Cho, D.I., Song, K.C., Noh, B.K.: Handbook of Analysis for Detection of Network Intrusion and Hacking. Infobook (2001)

[19] http://www.netfilter.org/, netfilter /iptables project homepage–The netfilter project

[20] ploug.eu.org/doc/s-netip.pdf

[21] http://www.netfilter.org/documentation/HOWTO/pt/packet-filtering-HOWTO.txt

[22] Peter Lichodzijewski, A.Nur Zincir-Heywood and Malcolm I. Heywood. Dynamic Intrusion Detection Using Self-Organizing Maps. CITSS, 2002

[23] The Third International Knowledge Discovery and Data Mining Tools Competition,http://kdd.ics.uci.edu/databases/kddcup99.kddcup99.html, May 2002.

[24] W. Lee, S. J. Stolfo and K. W. Mok, "Mining in a data-flow environment: experience in network intrusion detection," in *Knowledge Discovery and Data Mining*, pp. 114-124, 1999.

[25] Stefano Zanero. Improving Self Organizing Map Performance for Network Intrusion Detection, 2004

[26] Kunal Chadha and Abhinav Kumar, Thesis submitted as part of Network Forensics Project, Jaypee Institute of information Technology University, Noida.

.

Challenges in Managing Information Security From an Organization's Perspective

Patrick Kanyolo Ngumbi
School of Science and Engineering
Atlantic International University
Hawaii, USA
pkngumbi@yahoo.com

Abstract: This study used purposefully selected employees to fill self-administered unstructured questionnaires to provide information on aspects concerning information security at organizational level. The responses were subjected to non-probability analysis from which understanding of challenges encountered and subsequent impact were obtained. Six evaluation questions were used to gain insight into information security components. The study documented four categories of challenges encountered, possible outcomes of challenges and consequential impact. These results are beneficial to business end-users, information security managers, top and senior management in organizations.

Keywords: Information security management, organizational level, business information systems, challenges, outcome, impact

I. INTRODUCTION

Information is very valuable business asset and it requires being suitably protected [1]. Protecting this information requires implementing appropriate information security measures. Measures are necessary tools to avoid occurrence of incidences from attacks.

Information security is preservation of [1]: confidentiality to ensure information can be accessed by those authorized; integrity to safeguard information accuracy and completeness; and, availability to ensure authorized users have access to information and associated assets.

The goal of information security is to provide effective level of protection. To realize this level, an information security management is necessary. This context of "management" assumes the definition from Glossary of Commercial Real Estate Terms [2], that, "management is a job of planning, organizing, and controlling business enterprise". Through planning, organizing and controlling, effective information security is achievable.

Information security management is concerned with making information protection more effective. Further, protecting business information effectively demands understanding of challenges pertaining to managing information security. Studies reviewed following aspects of information security: (1) Lack of proactive actions on information security management [3], which means that organizations are ill-prepared for eventualities; (2) New and evolving technologies, research, tools and standards pose new challenges to organizations [4], which means it is a source of difficulties in securing business transactions, infrastructure and information; and, (3) Four challenges identified as structural, process, boundary and human, have challenges concerning human resources least emphasized despite having consequences in threats from inside organizations [5].

To advance understanding in the area of business information protection, this study examines challenges in information security management through organizations' employees. The study uses the research question: "What are today's organizational challenges constraining effective management of information security".

The understanding of challenges is beneficial to information security managers and decision makers in organizations. The study scope entails reviewing relevant literature on one hand and carrying out non-probability analysis of responses on the other hand, to obtain answer to the research question. Uses of results of this study include security managers determining threats and vulnerabilities in order to maintain effective risk management and enabling interlink for strategic, tactical and operational security levels.

II. RELEVANT WORK

2.1 Information Security Management

International Organization for Standardization (ISO) 17799 [1] provides three basic information security goals, namely, confidentiality, integrity and availability. To achieve the goals an organization needs to implement management and technical security measures. From management security measures, the organization can attain physical and operational security as well as legal and ethical obligations. On the other hand, from technical security measures an organization can attain following: access controls, system integrity, cryptography for security, audit and monitoring, and, configuration and security assurance.

Today's information security focus is to secure business information systems [6]. Further, today's business environment is complex and sometimes it involves real-time transactions, which can be prone to myriad of security attacks. This scenario necessitates a management approach which is information security management. Information security management is defined in Vermeulen and Von Solms [7] as "… the structured process for implementation and ongoing management of information security in an organization". It is a process that is structured – meaning, it is a prearranged set of procedures for information security to implement. It is also an ongoing management – meaning that, it is a continuous activity of planning, controlling, coordinating or organizing information security.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Components of information security management are: security objectives, business requirements, risk management, identity and access management, security policies and procedures, threats and vulnerabilities, security domain management, and incident response [8]. Security objectives involve confidentiality of information, integrity of information and availability of resources. Business requirements entail legal and operational requirements. Risk management involves balancing need for availability, integrity and confidentiality requirements vis-à-vis selection of safeguards for threats and vulnerabilities. Identity and access management ensures applications distinguish users from non-users and provide services appropriate to different users. Through security policies and procedures, security management on threats are identified and suitably implemented. Security domain management entails limiting threats and vulnerabilities of organization information. Incident response is a requirement that requires procedures to be in place to handle incidents as and when they occur.

Information security standards can be used to provide standard mechanisms to protect information. Standards are used to develop and benchmark security management programs. Information security standards are management standards used to guide top executives and senior managers through issues and to develop potentially effective information security management program. Details of information security standards are found in ISO/IEC 27001 [9] and ISO/IEC 27002 [10].

Today, business information requires more than just technology-centered security approach for it to be effectively managed. Kalkowska found individual and organizational values are important when it comes to effective information security management, and further that, it is difficult to formalize behavior of employees by only rules, procedures or even regulations [11]. Instead, to influence changes for information security one may need to target culture of organization as pointed out by Hofstede [12].

Top and senior management information security management concerns are found in three organizational security levels, namely, strategic, tactical and operational security levels [13]. Information requirements for security management are policy-driven at the strategic security level when management is guideline-driven at the tactical security level and measures-driven at operational security level. Further, strategic level issues affect organization strategy when tactical issues relate to processes and methodologies used in managing security; operational level installation and operation of security tools, and measures are prominent operations of organization [13]. A further aspect of information security is that it requires integration with other strategic parts of business to make senior management agenda [14, 15].

2.2 Information Security Governance

The need for information retention and privacy coupled with significant threats of information system disruptions from hackers, worms, viruses and terrorists have resulted in a need for a governance approach to protecting information and reputation. Drucker [16] stated that, "The diffusion of technology and the comodification of information transformed the role of information into a resource equal in importance to the traditionally important resource of land, labor and capital". Between then and now, this value escalated and dependence on information increased exponentially [17]. Further, a large portion of the task in protecting critical information resources falls squarely on shoulders of executives and boards [17].

Information security is a technical issue, business and governance challenge that involves adequate risk management, reporting and accountability [17]. An effective information security requires active involvement of executive so that tasks such as assessment of emerging threats and organization's response to them have corporate support. In order to have an effective information security governance, boards and senior executives must have following: a clear understanding of what to expect from the information security program and the need to know how to direct the implementation of program; how to evaluate their own status pertaining to existing program; and, how to decide on the strategy and objectives of an effective program [17].

Information security governance in essence involves leadership, organizational structures, and processes [17]. Information Technology Governance Institute (ITGI) [17] gives a summary for five basic outcomes of information security governance as:

1. Strategic alignment of information security with business strategy to support objectives.
2. Risk management by executing appropriate measures to manage and mitigate risks and reduce potential impacts on information resources to acceptable level.
3. Resource management through utilizing information security knowledge and efficient and effective infrastructure.
4. Performance measurements through measuring, monitoring and reporting information security governance metric to ensure that organization's objectives are achieved.
5. Value delivery by optimizing information security investments in support of organization's objectives.

III. RESEARCH THESIS AND APPROACH

In line with recommendations from Dick [18] that research question should be kept general, flexible and open with what is happening, this study's research question is: "What are today's organizational challenges constraining effective management of information security?" To focus and seek insight from components of information security aspects, the study used six evaluation questions as follows: (1) How organizations are affected by change of focus to securing business information systems; (2) What tools/security measures are in use for information security; (3) What processes/systems are in use to manage information security; (4) What mechanisms are implemented to protect against threats/prevent vulnerabilities; (5) What challenges are hindering effectiveness of information security management; and, (6) What the impact from challenges are.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Qualitative research approach was adopted in this study in accordance with Denzin and Lincoln [19] definition that qualitative research is the study of things in their natural settings aimed at making sense of or interpreting the meanings people bring to them. The study used research question to state and focus on the understanding being sought as recommended in Creswell [20].

Purposeful sampling selection was used to identify participants involved with either information security management or information security decision making. The sample represents an indefinite population because it is not possible to know the many organizations fitting this selection. There is, though, possibility of bias in this selection considering that not every potential selection has equal possibility of being selected. This study selection is small since the participants were fifty, which coupled with purposefully selected Information and Communication Technology (ICT) professionals, makes the sample tolerably reliable and adoptable with an added advantage that time and money were saved [21].

Fifty self-administered unstructured questionnaires were sent out and thirty two respondents returned theirs filled. This data became the primary data for qualitative analysis. Results of this analysis coupled with relevant literature review results provided study results. Interpretive research was adopted for data analysis, where the meaning follows from explanation in Walsham [22] that, it neither predefines dependent/independent variables nor sets out to test hypothesis but instead aims to produce understanding of social context of phenomenon and process. Further, according to Orlikowski and Baroudi [23], understanding social process involves getting inside the world of those generating it; hence the study used responses of employees to obtain insight into processes/systems in organizations.

Analysis was carried out as follows: (1) scrutinized questionnaires for accuracy and consistency; [2] identified and categorized main themes, topics or patterns; and, (3) interpreted by use of contents and commonalities coupled with relevant literature review to give answers to evaluation questions and consequently the research question of study.

IV. DISCUSSION AND RESULTS

4.1 Discussion of Responses

Discussion of evaluation questions follows below.

4.1.1 How organizations are affected by change of information security focus to securing business information systems

Figure 1 shows that majority of organizations reacted to change of focus by introducing new solutions commensurate with new challenges. These are: “New solutions for new challenges” which involve introducing new security tools and/or technologies, upgrading networks and/or systems, and, implementing security measures to guard against internal and external threats. Other measures taken but by fewer organizations are: “Awareness campaigns and/or skills development” and “Introducing an information security function”. Responses involving “Improved system administration” and “Focusing on effective team work and knowledgeable employees” were reported but appear they were not common reactions.

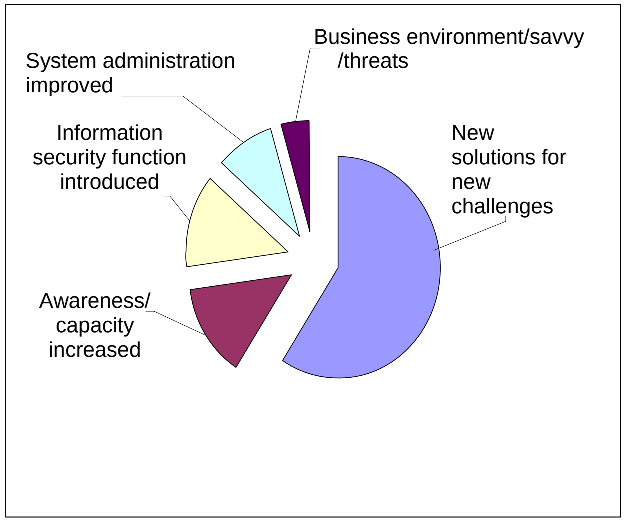


Figure 1: Organization’s reactions to information security change focus

Figure 2 shows responses on what was affected in information security focus change. Responses indicate that where information security focus changed there was pronounced change in internal/external user protection followed by change in the approach to information protection. Internal/external information user protection affects access controls and IT infrastructure technologies. These findings agree with what is expected considering that a change in approach to information protection would involve consideration/adoption of following measures: (1) Minimizing chances for malicious hackers to succeed, (2) Users getting no more privileges than necessary to do their job assignments, and, (3) Granting permissions to users based upon separation of privileges.

More organizations appear to have resulted to adopting new solutions for new challenges when few organizations appear to have emphasized awareness and/or skills development. Further, fewer organizations reacted by introducing information security function. These changes appear to have centered on upgrading or acquisition of technologies for assuring security for business information systems. It appears therefore that these organizations adopted new security measures coupled with new technology to provide sufficient protection in this new environment under question.

4.1.2 What security tools or measures are in place for managing information security

Table 1 shows the security tools/measures present in the organizations. At strategic security level, written security policies were reported as more commonly available than

others. Written security policies are followed by existence of security objectives and goals, which, since they are part of security policy, may be seen as confirming presence of security policy. The other tool/measure in the strategic security level but reported by fewer respondents, is security architecture, which indicates presence of documented designs on security. At the tactical security level, security procedures followed by security benchmarks and then standards are pronounced. The least pronounced tactical security tool/measure is the process methodology, which is an indication of lack of international certification. At operational security level, network, physical, data, application and infrastructure security measures exist.

Figure 3 provides responses on how organizations developed their security goals and objectives. Responses indicate that "Consultation within senior management, technical departments and other stakeholders" has the highest number of responses followed by "Assignment to persons to produce required critical data and resources for protection", and "Consultation within and inter departments and senior management". Use of "Adoption and ad hoc methods" and "Policies/strategic plans" were least reported.

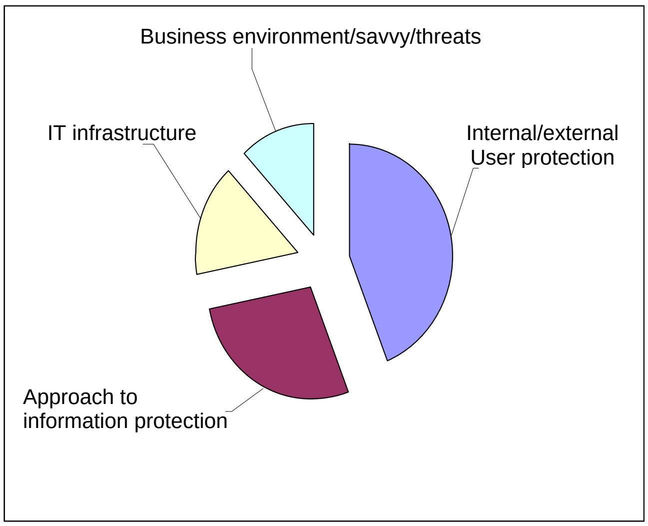


Figure 2: What was affected in the focus change in information security

The main security measure/tool used appears to be "Written security policies" at the strategic level while "Security procedures" form the measure/tool at the tactical security level. "Network security", "Physical security" and "Data security" form the common measures/tools at operational security level. The use of consultation between senior management, technical and other stakeholders appears common in developing security goals and objectives.

4.1.3 What are the processes or systems in use to manage information security in the organizations

Figure 4 shows responses on formal processes/systems used in organizations in managing information security. Responses show organizations used "Automated/written/ unwritten procedures" to manage information security, followed by use of "ICT policy guidelines". Reported but minor processes include: "IT manager prescribing" and "Being reactive to issues". Notable of these responses on this aspect is that some respondents believed their organizations did not have a process for managing information security and an information security management system appears least in existence.

Table 1: What information security tools/measures existed at different security levels

Security measures/tools		Security levels		
		Strategic	Tactical	Operational
1	Written security policies	22		
2	Security objectives	18		
3	Security goals	17		
4	Security architecture	8		
5	Security procedures		18	
6	Security benchmarks		9	
7	Standards		8	
8	Process methodology		5	
9	Network security measures			23
10	Physical security measures			21
11	Data security measures			20
12	Infrastructure measures			18
13	Application security measures			16
14	Disguise custody of equipment			1

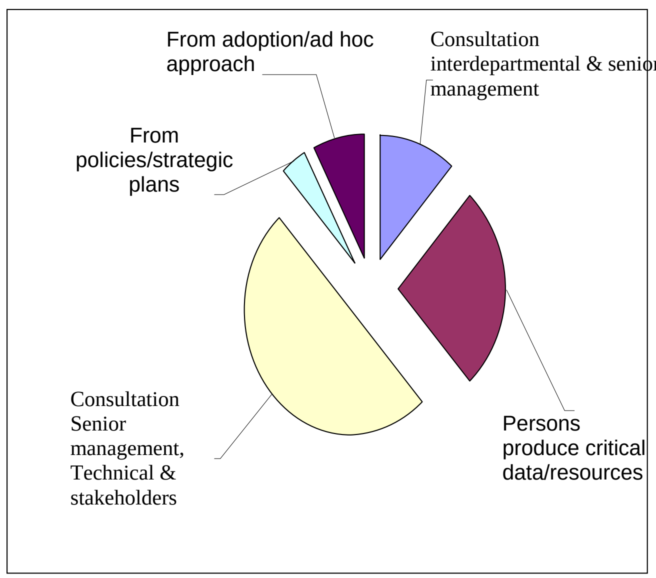


Figure 3: How security goals and objectives were developed

Procedures are common as process/system used in managing information security. Table 2 shows responses on

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

the processes used when checking individuals dealing with critical responsibilities. To check individuals, the process involved maintaining “Different and accountable roles with privileges and/or audits”, followed by “Staff vetting” and “performance contracting” in that order. “Regular surveys & reviews” were least reported as processes for checking individuals.

Processes or systems used in managing information security are thus automated procedures, written or unwritten procedures, and ICT policy guidelines. When dealing with critical assignments, employees are checked through maintaining different/accountable roles in assignments in addition to differing privileges and occasional audits. Individuals can be vetted and performance contracting is employed in some organizations though not common.

4.1.4 What mechanisms are implemented to protect against threats and prevent exploitation of vulnerabilities

Table 3 shows responses on the mechanisms used to protect organization technology, physical and logical access, applications and data. Ordered by number of responses reported against its use, mechanisms identified can be outlined as follows:

(a) Firewall policy. It protects information and systems against external and internal security threats.
(b) Access, password and antivirus policies. Responsible for preserving confidentiality, integrity and availability of information.
(c) Backups and business continuity programs. Responsible for ensuring continuity of services and availability of information.
(d) Physical security measures. These measures involve combining locks and guards to deter and ensure sensitive documents, business information systems, and servers, are not accessed by unauthorized persons.
(e) System administration roles. These measures ensure automated procedures and policies are not only implemented but are also monitored and reviewed accordingly.
(f) Intrusion Detection System (IDS) and Intrusion Protection Systems (IPS). The existence of these systems ensures organizations can identify and prevent harmful incidences to business information systems and further automatically log incidences for future learning and review.
(g) Encryption of data. Through this mechanism, organizations can have assurance in integrity and availability of its information and systems.

Therefore, the main mechanisms used to protect business information systems against internal and external threats involve implementing firewalls, access policies, password policies and antivirus policies. Backup and business continuity plans coupled with physical measures ensure continuity of business operations and availability. To a lesser extend, system administration, Intrusion Detection Systems/Intrusion Prevention Systems and encryptions are used.

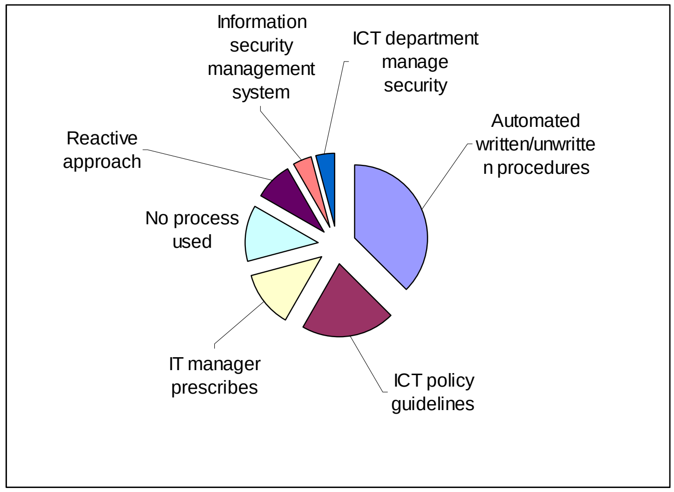


Figure 4: Formal processes used in managing information security in organizations

Table 2: Processes used to check individuals occupying critical positions in organizations

	Response	Implication	Number
1	Maintaining different and accountable roles and privileges and/or audits	Can detect tendencies and prevent internal threats. Can facilitate improved compliance hence have management with ethics, predictable outcomes and threat-management	6
2	Vetting through government machinery	Protects organization against criminal inclined employees hence internal threats minimized	3
3	Performance contracting and appraisal	Motivates, rewards and reprimands individual performance thereby cultivating a responsible positive culture	3
4	Regular surveys and constant reviews	Protects organization from internal attacks by constantly monitoring employee actions and tendencies	1

4.1.5 What challenges are responsible for hindering effectiveness of information security management in these organizations

Based on the received responses, four categories of challenges affecting organizations were identified as challenges encountered in ineffective information security, when integrating information security management function into other business processes, when identifying IT infrastructure, and, when securing IT infrastructure. Brief outline of each of the challenges follows below.

 http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Table 3: Mechanisms used to protect technology, physical and logical access, applications and data

	Response	Number
1	Firewall policy	10
2	Access and password policies	9
3	Antivirus policy	9
4	Backups & Business Continuity Programs	6
5	Security guards for critical areas	5
6	Use of combination locks	4
7	Surveillance cameras and alarm systems	3
8	System administration roles	3
9	Gate passes	2
10	Configuration policy	2
11	Intrusion detection and prevention systems	2
12	Encrypting data	2
13	Surveys and reviews	1
14	Copy rights	1
15	Motivate personnel	1
16	Rotation and/or separation of duties	1
17	Manager and staff affair	1

(a) Challenges encountered in information security duties

Table 4 provides challenges reported encountered when performing information security. It is because of these challenges that preventing unauthorized access, use, disclosure, disruption, modification or destruction of information and business systems is difficult or unachieved. Identified challenges under this category follow below.

1. Lacking information security management system. Lack of information security management system shall provide an ineffective protection.
2. Lacking or insufficient top/senior management support. This leads to inability to provide necessary protection to realize effective management.
3. Lacking or insufficient capacity, motivation or integrity for supporting and maintaining information security implementations. This leads to ineffective or compromised protection.
4. Lacking or insufficient up-to date awareness of threats to information security. This leads to ineffective protection in the organization.
5. Lacking or insufficient end-user information security awareness, skills development and understanding of their roles. This leads to lack of protection against internal threats and prevention of exploitation of vulnerabilities.
6. Dynamic technological changes. Such changes lead to inappropriate solutions and ineffective protection of information security.
7. Dynamism and complexity in information sharing and access. This situation makes it difficult to realize effective and sufficient protection in the organization.
8. Balancing need to know and be accessible. Through access, internal and external threats may be realized. When realized, the threats render protection less effective.

9. Procurement bureaucracies. Subsequent delays associated with this situation may compromise information and system protection.

These challenges bring about inability to protect business information systems. When attended, it leads to either implementing an effective system/mechanism or empowering/facilitating protection provision. If unattended, security attacks can succeed and vulnerabilities can easily be exploited. Lack of information security management system is the major challenge reported in this category.

(b) Challenges encountered when integrating information security management function to other business processes

Table 5 provides responses on challenges encountered when integrating information security management function with other businesses. Brief outline of challenges identified under this category follow below.

1. Lacking or insufficient ownership and understanding of the top management duty and role in supporting information security. Inadequate budgetary support and inappropriate acquisition for measures/tools are responsible.
2. Lacking or insufficient technical capacity. This situation makes it difficult to design, implement and maintain measures/tools for the protection.
3. Lacking or insufficient user awareness. This situation makes support to implement and maintain security measures/IT infrastructure inadequate.
4. Inappropriate or inadequate IT infrastructure. This situation leads to insecure business operations.
5. Cost taking precedence to acquisitions for security measures, tool or IT infrastructure in decisions. This leads to insecure business operations.
6. User lethargy. This makes it difficult to get adequate user support for continuous and efficient service delivery.
7. Faulty system requirements development. This leads to wrong designs and acquisitions for IT infrastructure and information security.
8. Lacking or poor IT governance. This leads to insufficient structures and capacity for managing IT infrastructure.
9. User apathy to changes. This leads to insufficient user support, ineffective operations and service delivery.
10. Lacking or insufficient inter-departmental communications. This leads to discontinuity of operations and inefficient service delivery.
11. Insufficient employee business support. This leads to discontinuous inefficient service delivery and vulnerable to internal threats.
12. Reliance to consultants. This can lead to possible compromise in confidentiality and integrity.
13. Lacking link between technical and management roles. This leads to discontinuity of operations and inefficient service delivery.
14. Lacking IT representation in strategic management levels. This can lead to insufficient understanding and support at the strategic level.

These challenges are encountered in projects involving planning and implementations. The outcome from this category of challenges is that protection measures will be based on inherent insecure implementation. Lack of ownership and understanding in top management, inadequate technical

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

capacity and lack of user awareness are the major challenges in this category.

Table 4: What are the challenges encountered when performing information security duties?

	Response	Number
1	Lacking or insufficient information security management system	11
2	Lacking or insufficient capacity, motivation or integrity for supporting and maintaining information security implementations	10
3	Lacking or insufficient top and/or senior management support	10
4	Lacking or insufficient up-to-date awareness of threats to information security	7
5	Technological change dynamism	5
6	Lacking or insufficient end-user information security awareness, skills and understanding of their roles	5
7	Dynamism and complexity in information sharing and access	2
8	Balancing the need to know and open information access	1
9	Costly security solutions	1
10	Procurement bureaucracies and/or subsequent delays	1

(c) Challenges encountered when identifying IT infrastructure

Table 6 provides responses on challenges reported as encountered when identifying IT infrastructure. A brief outline of challenges identified in this category follows below.

1. Faulty procurement/wrong solution provider. This leads to wrong solutions rendering discontinuity of operations and inefficient service delivery.
2. Inadequate technical involvement and knowledge. This leads to faulty or wrong solutions, acquisitions and implementation.
3. New technologies always emerging in ICT within very short time. This makes it difficult to identify appropriate solutions or even cope with changes.
4. Costly technological solutions vis-à-vis organizational growth. This makes it difficult for governance to sufficiently support relevant budget for procurement.
5. Lack of adequate user awareness of available technological solutions. This leads to insufficient user support and participation.
6. Lack of or insufficient top management awareness of technological solutions. This leads to faulty or wrong IT infrastructure solutions.
7. Increasing complexity of environment and platform. This makes it difficult to attain appropriate designs for IT infrastructure solutions.
8. Lack of or insufficient IT infrastructure alignment to service delivery. This leads to inefficient service delivery.

These challenges are encountered when identifying software for development, software for maintenance, software for purchase, IT hardware, and IT service delivery. If not attended, outcome involves insecure business operations and inefficient service delivery. Faulty procurement or wrong solution provider is the major challenge reported in this category.

Table 5: Challenges encountered when integrating information security management function with other business processes

	Response	Number
1	Lacking ownership and understanding by top management	5
2	Inadequate technical capacity	5
3	Lack of user awareness	5
4	Inappropriate infrastructure	3
5	Cost taking precedence at expense of acquired	3
6	User lethargy	2
7	Poor or faulty system requirements	2
8	Lack of IT governance	2
9	User apathy to change	2
10	Lack of or insufficient inter-departmental communication	2
11	Insufficient employee business support	1
12	Over reliance to consultants	1
13	Lack of link between technical and management roles	1
14	Lack of IT representation in strategic management levels	1

Table 6: What challenges are encountered when identifying IT infrastructure?

	Responses	Number
1	Faulty procurement or wrong solution provider	12
2	Inadequate technical involvement and knowledge	7
3	New technologies always emerging in ICT within very short time	5
4	Costly technological solutions vis-à-vis the organizational growth	5
5	Lack of adequate user awareness of available technological solutions	3
6	Lack of or insufficient top management awareness of technological solutions	2
7	Increasing complexity of environment and platform	2
8	Lack of or insufficient IT infrastructure alignment to service delivery	1

(d) Challenges encountered when securing IT infrastructure

Table 7 provides responses reported as challenges encountered when securing IT infrastructure. Brief outline of the challenges follows below.

1. Lack of or insufficient skills. This leads to inadequacy in facilitating and supporting security solutions.
2. Bureaucracy and unstructured approach to acquisition of security solutions. This leads to delay and faulty security solutions.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

3. Lack of or insufficient awareness of threats in all stakeholders. This leads to insufficient support and participation in implementing security solutions and avoiding risks.
4. Inadequate access control measures. This leads to possibilities of unauthorized access, disclosure and alteration.
5. Growing sophistication and diversification of attacks. This leads to lack of protection against unknown threats and vulnerabilities.
6. Lack of or insufficient support by top management. This leads to insufficient support to budgetary allocations for security solutions.
7. Lack of or insufficient measures and policies to combat threats. This leads to inadequate plans and protection.

Table 7: What are the challenges in securing IT infrastructure?

	Responses	Number
1	Lack of sufficient skills	9
2	Bureaucracy/unstructured security solutions acquisition	5
3	Insufficient awareness of threats in stakeholders	5
4	Inadequate access control measures	4
5	Sophistication and diversification of attacks	3
6	Lack of sufficient support by top management	3
7	Lack of sufficient measures and policies	1
8	Growing volumes in transactions	1
9	Internal threats and insecure systems	1
10	Costly security solutions	1
11	Lacking mechanisms to sufficiently mitigate risk in outsourcing	1
12	Being limited in technological solutions	1

8. Growing volumes in transactions. This leads to varying solutions for storage and transmissions at the organizational level.
9. Internal threats and insecure systems. This leads to vulnerable business information systems.
10. Costly security solutions. This leads to inadequate protection.
11. Lacking or insufficient mechanisms to mitigate risk in outsourcing.
12. Being limited in technological solutions. This leads to inadequate designs and solutions in the protection of business information systems.

These challenges are encountered when securing software development, software maintenance, software purchase, IT hardware, and service delivery. If not attended, the outcome will be insecure IT infrastructure and operations. Lack of sufficient skills is the major challenge in this category.

4.1.6 What is the impact from challenges responsible for inhibiting effectiveness of information security management

Table 8 shows the possible outcome and eventual impact from identified challenges. Thirteen possible outcomes were identified from which eventual impact is possible. Outline follows below.

(a) Insufficient protection is caused by lack of information security management system, management support or existence of internal threats.
(b) Insufficient support and participation are brought about by lack of sufficient capacity, motivation, security awareness, internal communications, support, policies and ownership.
(c) Inability to cope can come from dynamic technological changes.
(d) Inadequate protection can come from dynamism and complexity found in information sharing and access.
(e) Over protection or under protection can come from a situation where balance for the need to know and comply with access needs is inadequately done.
(f) Delays in acquiring security solutions can come from bureaucratic and unstructured methods found in the acquisition of security solutions.
(g) Ineffective service delivery can come from inappropriate or inadequate IT infrastructure, lack of IT governance or lack of IT infrastructure alignment to service delivery.
(h) Lacking appropriate security solutions possible if cost of security infrastructure affects decisions during acquisition which lead to inappropriate security solutions.
(i) Inappropriate security solutions can result from growing sophistication and diversification of attacks.
(j) Faulty security solutions can originate from use of faulty system requirements, which lead to faulty security designs.
(k) Insufficient protection can be caused by costly security solutions which influence decisions responsible for insufficient information protection.
(l) Compromised confidentiality and integrity is possible from over reliance to consultants or lack of mechanisms to mitigate risks in outsourcing.
(m) Insufficient data security can come from growing volumes in transactions common nowadays.

4.2 Results of Study

Table 9 is a summary of results from the evaluation questions in the study. The table provides specific results of the six evaluation questions used.

The following can be said about the organizations sampled:

4.2.1 That, organizations appear to have reacted to change of focus to securing business information systems by adopting new security measures together with acquisition of relevant technologies.
4.2.2 That, written security policies, security procedures, network security measures, physical security measures, and data security measures were major tools used to manage information security.
4.2.3 That, processes in use involve automated procedures, written/unwritten procedures and ICT policy guidelines.
4.2.4 That, mechanisms used to realize security involve implementing firewalls, access policies, password

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

policies and antivirus policies together with backup and physical measures.

Table8: Possible outcome and eventual impact from identified challenges

Identified challenges		Possible challenge outcome	Eventual impact
1	Lacking information security management system, management support and existence of internal threats/insecure systems	Insufficient protection	Loss of capital, reputation or even business opportunities
2	Lacking/insufficient capacity, motivation, security awareness, internal communications, support, policies and ownership	Insufficient support and participation	
3	Dynamic technological changes	Inability to cope	
4	Dynamism and complexity in information sharing and access	Inadequate protection	
5	Balancing need to know and compliance to being open	Over protection or under protection	
6	Bureaucratic and unstructured methods in acquisitioning security solutions	Delays and insufficient protection	
7	Inappropriate or inadequate IT infrastructure, lacking IT governance and IT infrastructure alignment to service delivery	Ineffective service delivery or business operations	
8	Consideration of cost at expense of acquisition of security infrastructure	Lacking appropriate security solutions	
9	Growing sophistication and diversification of attacks	Inappropriate security solutions	
10	Faulty system requirements	Faulty security solutions	
11	Costly security solutions	Insufficient protection	
12	Over reliance to consultants or insufficient mechanisms to mitigate risk in outsourcing	Compromised confidentiality and integrity	
13	Growing volumes in transactions	Insufficient data security	

4.2.5 That, effective management of information security is hindered by challenges encountered in integrating information security management function to other businesses, in identifying IT infrastructure, in securing IT infrastructure, and in the program for information security program.

4.2.6 That, information security assurance or lack of it depends on the acquisition and implementation of security solutions, business operations, and management aspects involved in protecting the business information systems.

4.2.7 That, the practice of using an information security management system was lacking in majority of organizations.

4.2.8 That, lack of ownership and understanding in top management, inadequate technical capacity and lack of user awareness was major problems in the organizations.

4.2.9 That, faulty procurement or wrong solution provider are problem when identifying IT infrastructure.

4.2.10 That, lack of sufficient skills in organizations is a major setback to security in an organization.

Where and when identified challenges are not mitigated, the result is ineffective information security management characterized by lacking protection to business information systems and eventual negative impact to business.

Table 9: Summary of results from study evaluation questions

	Evaluation question	Results
1	How the change of focus to securing business information systems was affected?	Organizations adopted new security measures coupled with new technology to provide sufficient protection. Awareness and skills development are least emphasized.
2	What security tools/ measures are in place to manage information security?	Organizations appear to have: (1) written security policies at the strategic level; (2) security procedures at the tactical level; and, (3) network, physical and data security measures at the operational level.
3	What processes/ systems are in use to manage information security	Automated procedures, written/unwritten procedures and ICT policy guidelines are the processes/systems used to manage information security. Employees are checked through maintaining different/accountable roles in assignments in addition to ensuring privileges and audits are employed..
4	What mechanisms are implemented to protect against threats and prevent exploiting vulnerabilities?	Organizations use implementations of firewalls, access policies, password policies and antivirus policies to protect business information systems against internal and external threats. Backup and business continuity plans coupled with physical measures ensure continuity of business operations and availability.
5	What challenges are responsible for hindering effectiveness of information security management?	Four categories of challenges identified are: (1) challenges encountered in protecting business information systems, (2) challenges encountered in integrating information security management function to other businesses, (3) challenges encountered when identifying IT infrastructure for business, and, (4) challenges encountered when securing IT infrastructure.
6	What is the impact from the identified challenges	Business information systems security attacks may be caused, enabled or facilitated by: (1) lack of, insufficient, compromised or ineffective protection, (2) faulty, wrong, insufficient or delayed security solutions, (3) inefficient and insecure business operations, and, (4) faulty, wrong and incomplete security solutions. The impact to business eventually is loss of capital, reputation and business opportunities.

V. CONCLUSION

The study identified four categories of challenges encountered in organizational management of information security. The study identified that there are challenges encountered when performing information security duties, integrating information security management function with

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

other business processes, identifying IT infrastructure and securing IT infrastructure.

Where the identified challenges are not mitigated accordingly, the result is ineffective information security management. The organization will experience lack of protection to business information systems and eventual negative impact to its business, which can translate into lost opportunities, reputation and capital. Organization will lack competitiveness and may even go under as a result.

This study has successfully obtained understanding of challenges in information security management from an organization's perspective as found today. The insight provides understanding of what system end-users, security managers and top/senior management should know and act on to realize effective management in organizational information security.

REFERENCES

[1] International Organization for Standardization/ International Electrotechnical Commission (ISO/IEC) 17799. Information Technology – Code of Practice for Information Security Management, International Standards Organization, 2000.

[2] Glossary of Commercial Real Estate Terms. Calgary Real Estate Board. Retrieved 5 April 2010. <http://www.creb.com/public/commercial-resources/ glossary-of-terms.php>

[3] R. C. Mitchel, R. Marcella and G. Baxter. Corporate Information Security Management, New Library World, Volume 100, Issue 5, 1999, 213 – 227.

[4] ISACA. An Introduction to the Business Model for Information Security, 2009. <www.isaca.org>.

[5] D. Ashender. Information Security management: A Human Challenge? Information security Technical Report, Volume 13, Issue 4, November 2008, 195-201, 2008.

[6] A. L. Nnolim and A. L. Steenkamp. Implementing a Planning Model for Information Security Management, International Journal of Computers, Systems and Signals, Volume 9, Number 2, 40-57, 2008.

[7] C. Vermeulen and R. Von Solms. The Information Security Management Toolbox – Taking the Pain out of Security Management", Information Management and Computer Security, Volume 10, Number 3, 119-125, 2002.

[8] Dan Sullivan. The Definitive Guide to Security Management, Realtime Publishers.com, 2006. <www.partnerprograminfo.com>

[9] ISO/IEC 27001. Information Security Management – Specification with Guidance for Use, International Standards Organization (ISO), 2000.

[10] ISO/IEC 27002 (2005). Information Technology – Code of Practice for Information Security Management, International Standards Organization (ISO), 2005.

[11] E. Kalkowska. Value Sensitive Approach to IS Security – a Socio-organizational Perspective, proceedings of the Eleventh Americas Conference on Information Systems, 2005.

[12] G. Hofstede. Measuring Organizational Cultures: A Qualitative and Quantitative Study across Twenty Cases, Administrative Science Quarterly, 35, 2, 286-316, 1990.

[13] P. Belsis, S. Kokolakis and E. Kiountouzis. Information Systems Security from a Knowledge Management Perspective, Information Management and Computer Security, Volume 13, November 3, 189-202, 2005.

[14] J. Wylder. Strategic Information Security, Auerbach/CRC Press LLC, 2004.

[15] V.Leveque. Information Security – A Strategic Approach, John Wiley & Sons, 2006.

[16] Peter Drucker. Management for the 21st Century, Harpers Business, 1993.

[17] Information and Communication Technology (ITGI). Information Security Governance: Guidance for Board of Directors and Executive management, 2nd Edition, 2006.

[18] B. Dick (2002). Grounded Theory: A Thumbnail Sketch, 2002. Viewed 1 February 2008. <http://www.scu.edu.au/schools/gcm/ar/arp/groundded.html>

[19] N. K. Denzin and Y. Lincoln. Introduction: The Discipline and Practice of Qualitative Research. Handbook of Qualitative Research, 2nd Ed. Thousand Oaks, CA: Sage, 2000.

[20] J. W. Creswell. Research Design: Qualitative, Quantitative and mixed Methods Approaches. Thousand Oaks, CA: sage, 2003.

[21] C. R. Kothari. Research Methodology: Methods and Techniques, 2nd Ed. New Delhi: New Age International Limited Publishers, 2004.

[22] G. Walsham. Interpretive Case Studies in IS Research: Nature interpretive, data analysis method and Method, European Journal of Information Systems, Volume 4, No. 2, pp. 74-81, 1995.

[23] W. J. Orlikowski and J. J. Baroudi. Studying Information Technology in Organizations: Research Approaches and Assumptions, Information Systems Research, 2(1): 1-8, 1991.

ACKNOWLEDGEMENT: The author would like to thank the Atlantic International University for support and partial scholarship which enabled completion of the thesis research, part of which is this paper. .

PROFILE: Patrick Kanyolo Ngumbi is a senior System Analyst in the National Social Security Fund in Kenya, charged with managing Data center. He presented this study results successfully in April 2010 for his final thesis to the Academic Department of the School of Science and Engineering, Atlantic International University for the degree of Doctor of Philosophy. He received his M.S. degree in Atmospheric Science from University of Wyoming (UW), USA in 1991 and B.Sc. (Honors) degree in Mathematics/Meteorology from University of Nairobi, Kenya in 1981.

Image Retrieval with Texture Features Extracted using Kekre's Median Codebook Generation of Vector Quantization

Dr. H.B.Kekre
Sr. Professor, MPSTME,
NMIMS Vileparle(W),
Mumbai 400056, India
hbkekre@yahoo.com

Sudeep D. Thepade
Ph.D. Scholar &
Assistant Professor,
MPSTME, NMIMS
Vileparle(W), Mumbai
400-056, India
sudeepthepade@gamil.com

Tanuja K. Sarode
Ph.D. Scholar MPSTME,
NMIMS Assistant
Professor, TSEC,
Mumbai 400-050, India
tanuja_0123@yahoo.com

Vaishali Suryavanshi.
Lecturer,
Thadomal Shahani Engg.
College, Bandra (w),
Mumbai 400-050, India
Vaishali.surya@gmail.com

***Abstract*—In this paper novel methods for image retrieval based on texture feature extraction using Vector Quantization (VQ) are proposed. We have used Linde-Buzo-Gray (LBG), and Kekre's Median Codebook Generation (KMCG) algorithms for texture feature extraction. The image is first divided into blocks of size 2x2 pixels (each pixel with red, green and blue component). A training vector of dimensions 12 is created using this block. Collection of all such training vectors is a training set. To generate the texture feature vector of the image, LBG and KMCG algorithms are applied on the initial training set to obtain codebooks of size 16, 32, 64, 128, 256 and 512. These codebooks are considered as feature vectors for CBIR. Thus the codebook generation algorithms and five different codebook sizes per algorithm result in 12 proposed image retrieval techniques. The proposed image retrieval techniques are tested on generic image database respectively having 1000 images. Results are also compared with the Gray Level Co-occurance Matrix (GLCM) method. The proposed CBIR methods outperform GLCM with higher precision and recall values. KMCG based CBIR give performance improvement over LBG based CBIR. The performance of KMCG CBIR improves with increasing codebook size. Overall in all KMCG CBIR with codebook size 512 gives best results with higher precision and recall values for both databases.**

Keywords— CBIR, Vector Quantization, GLCM, LBG, KMCG

I. INTRODUCTION *(HEADING 1)*

Technological advances in digital imaging, broadband networking, and data storage have motivated people to communicate and express by sharing images, video, and other forms of media online [1,3]. Although the problems of acquiring, storing and transmitting the images are well addressed, capabilities to manipulate, index, sort, filter, summarize, or search through image database lack maturity [31]. Modern image search engines [31, 37] retrieve the images based on their visual contents, commonly referred to as Content Based Image Retrieval (CBIR) systems [40]. CBIR systems have found applications in various fields like fabric and fashion design, interior design as panoramic views [17,18,32-35], art galleries [32], museums, architecture/engineering design [32], weather forecast, geographical information systems, remote sensing and management of earth resources [38,39], scientific database management, medical imaging, trademark and copyright database management, the military, law enforcement and criminal investigations [25], intellectual property, picture archiving and communication systems, retailing and image search on the Internet. Typical CBIR systems can organize and retrieve images automatically by extracting some features such as color, texture, shape from images and looking for similar images which have similar feature [36, 37]. CBIR systems operate in two phases. In the first phase, feature extraction (FE), a set of features, called feature vector, is generated to accurately represent the content of each image in the database. A feature vector is much smaller in size than the original image [29, 30]. In the second phase, similarity measurement (SM), searching distance between the query image and each image in the database using their signatures is computed so that the most similar images can be retrieved [24,28]. A variety of feature extraction techniques have been developed. Color based feature extraction techniques include color histogram, color coherence vector, color moments,, circular ring histogram [4], BTC extensions [25, 28, 30]. Texture based feature extraction techniques such as co-occurance matrix [6], Fractals [5], Gabor filters [5], variations of wavelet transform [1], Kekre transform [17, 27, 39] have been widely used. Effort has been made in even to extend image retrieval methodologies using combination of color and texture as the case in [23] where Walshlet Pyramids are introduced. The synergy resulting from the combination of color and texture is demonstrated to be superior than using just color and texture [37, 38].

In section II texture feature extraction using GLCM and VQ based methods viz. LBG and KMCG are discussed. In section III, technique for image retrieval using vector quantization is proposed. Results and discussion are given in section IV and conclusions are presented in section V.

II. TEXTURE FEATURE EXTRACTION METHODS

Texture is important component of human visual perception and can be effectively used for identifying different image

regions [1]. Compared with color and shape features, texture features indicate the shape distribution, better suits the macrostructure and microstructure of the images [5]. Texture representation methods can be classified into three categories, namely structural, statistical and multi-resolution filtering methods. The identification of specific textures in an image is achieved primarily by modeling texture as a two-dimensional gray level variation [6,37]. This two dimensional array is called as Gray level Co-occurance Matrix (GLCM). GLCM describes the frequency of one gray tone appearing in a specified spatial linear relationship with another gray tone, within the area under investigation.

A. *GLCM Method*

Normalized probability density Pij of the co-occurance matrices can be defined as follows:

$$P_\delta(i,j) = \frac{\#\{|(x,y),(x+d,y+d)| \in S \mid f(x,y)=i, f(x+d,y+d)=j\}}{\#S} \quad (1)$$

where x, y = 0,1,…..L-1 are the gray levels. S is set of pixel pairs which have certain relationship in the image. #S is the number of elements in S. $P_\delta(i,j)$ is the probability density that the first pixel has intensity value i and the second j, which are seperated by distance δ=(dx, dy).

The GLCM is computed in four directions for $\delta=0^0$, $\delta=45^0$, $\delta=90^0$, $\delta=135^0$. Based on the GLCM four statistical parameters energy, contrast, entropy and correlation are computed. Finally a feature vector is computed using the means and variances of all the parameters [8, 9]. The steps for texture feature extraction using GLCM are as given below

1. Separate the R,G, B planes of image.
2. Repeat steps 3-6 for each plane.
3. Compute four GLCM matrices (directions for $\delta=0^0$, $\delta=45^0$, $\delta=90^0$, $\delta=135^0$) as given by eq. (1)
4. For each GLCM matrix compute the statistical features Energy(Angular second moment), Entropy(ENT), Correlation(COR), Contrast(CON) [8,9] using the equations mentioned below:

   Energy: measures textural uniformity (i.e. pixel pairs repetitions).and can be given as Angular Second Moment (ASM)

$$ASM = \sum\sum P^2(i,j) \quad (2)$$

   Contrast (CON): Contrast indicates the variance of the gray level .

$$CON = \sum\sum (i-j)^2 P(i,j) \quad (3)$$

   Entropy (ENT) : This parameter measures the disorder of the image. For texturally uniform image, entropy is small.

$$ENT = -\sum\sum P(i,j)\log[P(i,j)] \quad (4)$$

   Correlation: (COR)

$$COR = \frac{\sum\sum ijP(i-j) - \mu_x\mu_y}{\sigma_x\sigma_y} \quad (5)$$

   Where μ_x, μ_y, σ_x, σ_y are the means and standard deviations of Px and Py respectively. Px is the sum of each row in co-occurance matrix. Py is the sum of each column in the co-occurance matrix.

Thus we obtained

ASM0	ENT0	COR0	CON0
ASM 45	ENT 45	COR45	CON45
ASM 90	ENT 90	COR90	CON90
ASM 135	ENT 135	COR135	CON135

Compute the feature vector using the means and variances of all the parameters. Thus, the feature vector f={μASM , μENT , μCOR , μCON , σASM ,σENT , σCOR , σCON } Where μ is mean and σ is variance of the parameters.

B. *VQ based methods*

Vector Quantization (VQ) [7-15] is an efficient technique for data compression [23]. VQ has been very popular in variety of research fields such as video-based event detection [26], image segmentation [19-22], speech data compression [23], CBIR [26,37,38] and face recognition [25].

VQ can be defined as the mapping function that maps k-dimensional vector space to the finite set **CB = { C1, C2, C3, ..., CN}**. The set CB is called codebook consisting of N number of codevectors and each codevector Ci = {ci1, ci2, ci3, ……, cik} is of dimension k. The key to VQ is the good codebook. This codebook is the signature/feature vector of the entire image and can be generated by employing various clustering techniques. The method most commonly used to generate codebook is the Linde-Buzo-Gray (LBG) algorithm [8]. The drawback of LBG algorithm is that the cluster elongation is -45^0 to 135^0 horizontal axis in two dimensional cases. This results in inefficient clustering. Kekre's Proportionate algorithm (KPE) removes the disadvantage of LBG [36]. However, LBG and KPE algorithms require heavy computations. Kekre's Fast Codebook Generation algorithm (KFCG) [12,24,36] requires less errors and least time to generate codebook as compared to other algorithms, as it does not require computation of Euclidian distance [2,33].

To generate the codebook, the image is first divided into fixed size blocks, each forming a training vector **X_i = (x_{i1}, x_{i2}, ……, x_{ik})**.The set of training vectors is a training set. This training set is initial cluster. The clustering algorithms like LBG, KPE, and KFCG etc are then applied on this initial cluster to generate the codebook of desired size. Below, LBG, KPE and KFCG algorithms for codebook generation are discussed.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

C. LBG Algorithm

In this algorithm centroid is computed as the first codevector for the training set. Two vectors v_1 & v_2 are then generated by adding constant error to the codevector. Euclidean distances [2,33] as presented in equation 6, of all the training vectors are computed with vectors v_1 & v_2 and two clusters are formed based on nearest of v_1 or v_2. This procedure is repeated for every cluster.

$$ED = \sqrt{\sum_{i=1}^{n} \left(X_{pi} - v_{qi}\right)^2} \qquad (6)$$

where, X_{pi} and v_{qi} are the training vector and codevectors respectively with size n.

D. Kekre's Median Codebook Generation algorithm (KMCG)

The steps of KMCG algorithm can be explained as follows.

- Image is divided into the windows of size 2x2 pixels (each pixel consisting of red, green and blue components).
- These are put in a row to get 12 values per vector. Collection of these vectors is a training set.
- The training set is sorted with respect to first column. The Median of the first column is used to divide the training set in two parts and the median vector is put in the codebook. Set the codebook size equal to 1.
- Further each part is then separately sorted with respect to second column to get two median values and these two median vectors are put into the codebook. Set the codebook size equal to 2.
- The process of sorting is repeated till codebook of desire size is obtained.

Here quick sort algorithm is used. This algorithm takes least time to generate codebook, since Euclidean distance computation is not required.

III. IMAGE RETRIEVAL USING VQ BASED TECHNIQUES

Image retrieval based on content requires extraction of features of the image, matching these features with the features of the images in the database and retrieving the images with the most similar features. Here, paper discusses the feature extraction technique based on vector quantization.

A. Proposed Feature Extraction Technique

1. Repeat steps 2-6 for each image in the image database.
2. Divide the image into blocks of size 2x2 (Each pixel having red, blue and green component, thus resulting in a vector of 12components per block)
3. Form the training set/ initial cluster from these vectors.
4. Compute the initial centroid of the cluster.
5. Obtain the codebook of desired size using LBG/KMCG algorithm. This codebook represents the feature vector/signature of the image.
6. Store the feature vector obtained in step 5 in the feature vector database.

B. Query Execution

For a given query image compute the feature vector using the proposed feature extraction technique. To retrieve the most similar images, compare the query feature vector with the feature vectors in database. This is done by computing the distance between the query feature vectors with those in feature vector database. Euclidian distance as given in equation (6) and correlation coefficient are most commonly used as similarity measure in CBIR [2, 36]. Here Euclidian distance is used as a similarity measure. As compared to GLCM method, this proposed method saves tremendous number of computations. Also the accuracy of the proposed VQ technique is much better than that of GLCM.

IV. RESULTS AND DISCUSSIONS

The proposed CBIR techniques are implemented in Matlab 7.0 on Intel Core 2 Duo Processor T8100, 2.1 GHz, 2 GB RAM machine to obtain results. The sample image database is shown in Figure 1. The results are obtained on the general database consisting of 1000 images from 11 different categories (some of these are taken from [41]). To test the proposed method, from every class five query images are selected randomly. So in all 55 query images are used from general database. To check the performance of proposed technique we have used precision and recall. The standard definitions of these two measures are given by following equations.

$$\text{Precision} = \frac{\text{Number\_of\_relevant\_images\_retrieved}}{\text{Total\_number\_of\_images\_retrieved}} \qquad (7)$$

$$\text{Recall} = \frac{\text{Number\_of\_relevant\_images\_retrieved}}{\text{Total\_number\_of\_relevent\_images\_in\_database}} \qquad (8)$$

The crossover point of precision and recall acts as performance measure of CBIR technique. Higher value of precision-recall at crossover point indicates better performance of image retrieval method. For the image database, results are obtained using LBG and KMCG for the codebook of sizes 16x12, 32x12, 64x12, 128x12, 256x12 and 512x12.

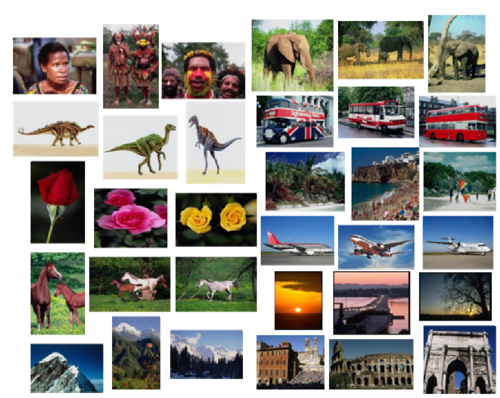


Figure 1 sample database of 11 images by randomly selecting one image from each category from general image database

Figure 2 shows average precision and recall plotted against number of retrieved images for General image database with precision-recall crossover point value 0.273.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

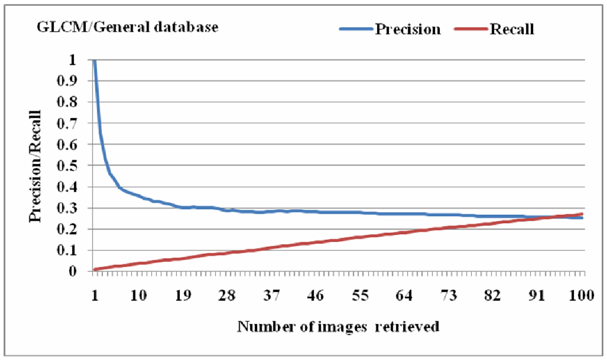


Figure 2. Average Precision and Recall plotted against number of retrieved images for GLCM-CBIR of General image database

Figure 3 gives crossover points of average precision and average recall values of the LBG based CBIR techniques for all codebook sizes tested on generic image database. Here the codebook sizes 32 and 64 are better with highest crossover point value. The precision and recall curves of codebook size 32 are higher than other codebook sizes indicating better performance for LBG-CBIR. The crossover points of average precision and average recall values for KMCG based CBIR techniques with considered codebook sizes are shown in figure 4. Here codebook size 512 is giving best performance.

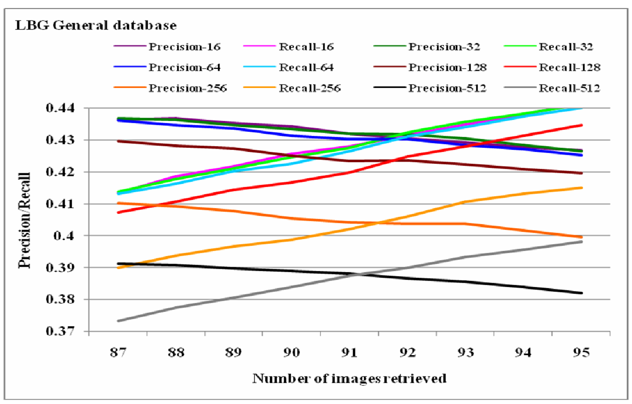


Figure 3. Average Precision and Recall plotted against number of retrieved images for LBG-CBIR of General image database

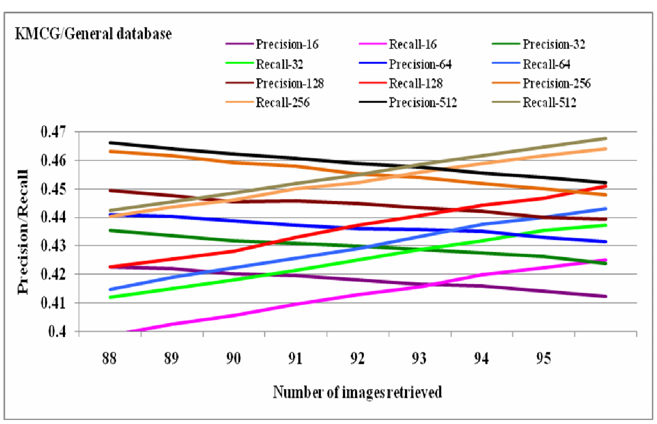


Figure 4. Cross-over points of average precision and recall using KMCG-CBIR for all considered codebook sizes of General image database

Figure 5 to figure 10 gives the comparison of crossover points of average precision and average recall values using LBG-CBIR and KMCG-CBIR for codebook sizes 16 to 512 respectively. In all codebook sizes KMCG-CBIR outperforms the LBG-CBIR with higher precision and recall values.

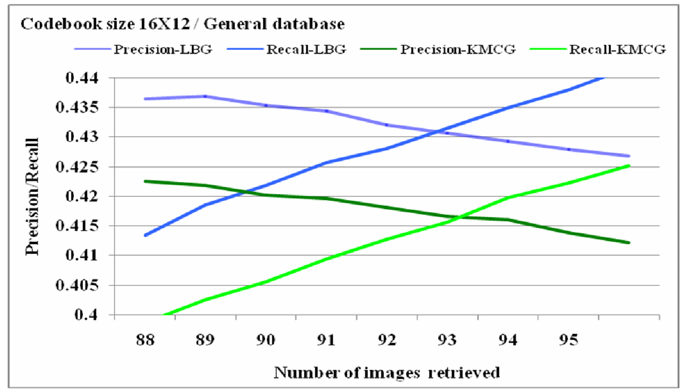


Figure 5. Comparison of Cross-over points of average precision and recall using LBG-CBIR and KMCG-CBIR for codebook size 16 of Generalimage database

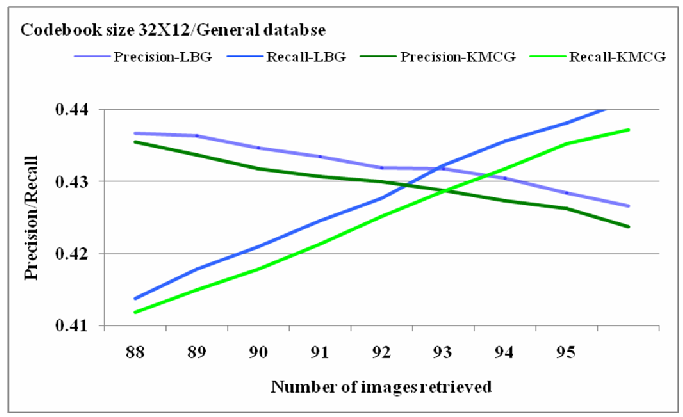


Figure 6. Comparison of Cross-over points of average precision and recall using LBG-CBIR and KMCG-CBIR for codebook size 32 of Generalimage database

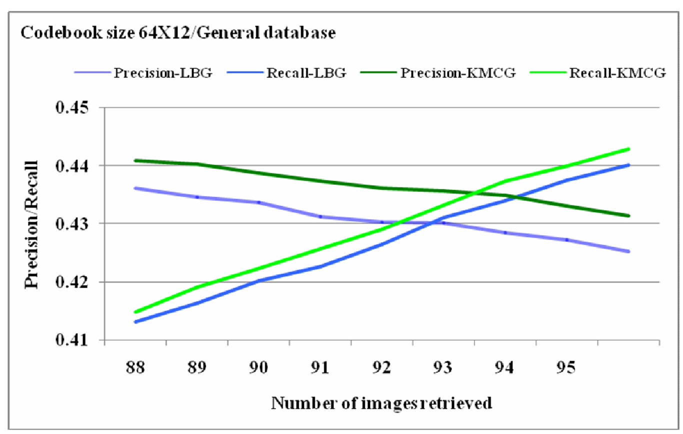


Figure 7. Comparison of Cross-over points of average precision and recall using LBG-CBIR and KMCG-CBIR for codebook size 64 of Generalimage database

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

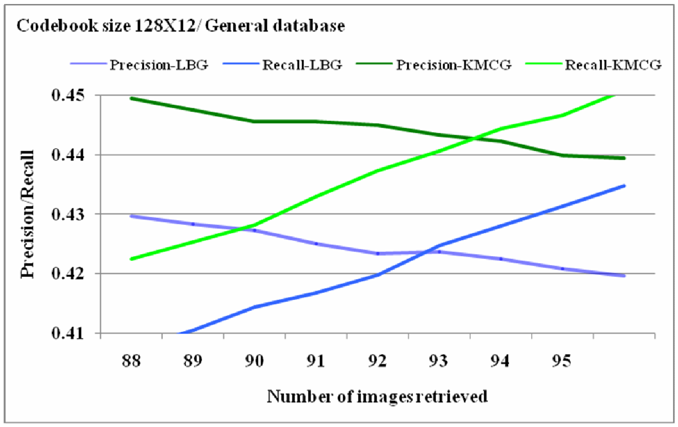


Figure 8. Comparison of Cross-over points of average precision and recall using LBG-CBIR and KMCG-CBIR for codebook size 128 of Generalimage database

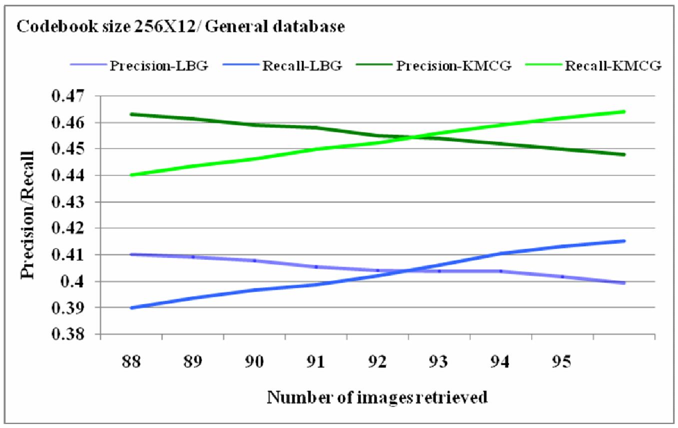


Figure 9. Comparison of Cross-over points of average precision and recall using LBG-CBIR and KMCG-CBIR for codebook size 128 of Generalimage database

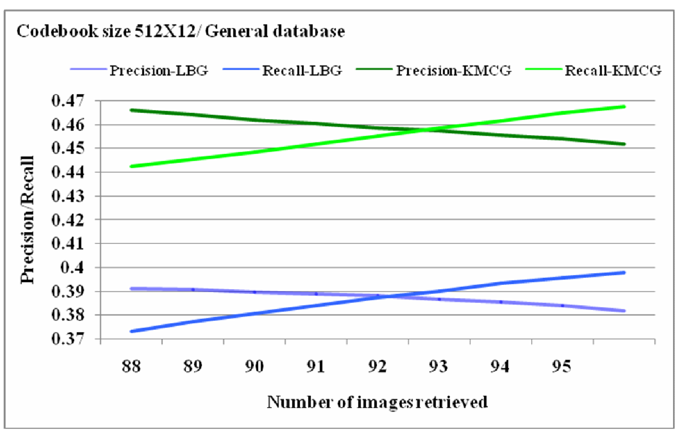


Figure 10. Comparison of Cross-over points of average precision and recall using LBG-CBIR and KMCG-CBIR for codebook size 512 of Generalimage database

Figure 11 gives comparison of average precision and average recall values of GLCM-CBIR with codebook generation based CBIR techniques with codebook size 512x12. Here the codebook generation based CBIR techniques are performing far better than GLCM-CBIR as indicated by higher average precision and recall values. Also the fact that KMCG-CBIR performs better than LBG-CBIR is reflected in the figure with higher precision nad recall values.

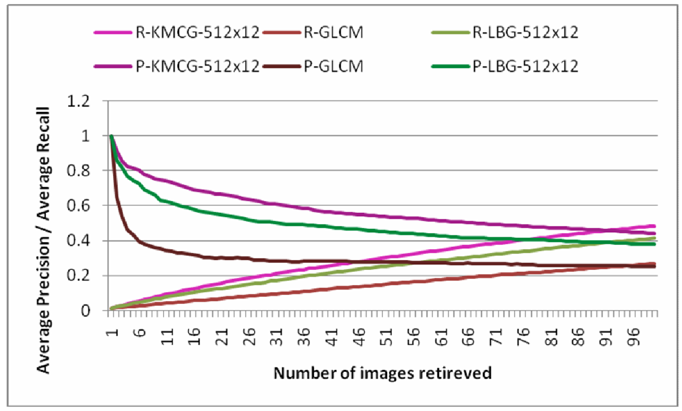


Figure 11. Comparison of average precision and recall using GLCM-CBIR, LBG-CBIR and KMCG-CBIR for codebook size 512 of Generalimage database

V Conclusion

The use of vector quantization codebooks as feature vectors for image retrieval is revisited in the paper. The paper used various codebook generation techniques such as of Linde-Buzo-Gray (LBG) and newly proposed Kekre's Median Codebook Generation (KMCG) algorithms for texture feature extraction. These codebooks extracted in sizes 16,32,64,128, 256 and 512 are used in proposed CBIR techniques. Thus two codebook generation algorithms and six different codebook sizes per algorithm result in 12 proposed image retrieval techniques. Results of the proposed CBIR techniques are also compared with the Gray Level Co-occurance Matrix (GLCM) method. The proposed CBIR methods outperform GLCM with higher precision and recall values, indicating better image retrieval. KMCG based CBIR give immense improvement over LBG based CBIR. In KMCG based CBIR the higher codebook size give better performance, proving the codebook size 512 to be the best. Overall in all codebook sizes KMCG gives best results with higher precision and recall values for both generic image database and COIL image database. The KMCG takes least time to form codebook (texture feature set) as compared to LBG. This proves that KMCG-CBIR gives not only better but also faster image retrieval.

REFERENCES

[1] Sanjoy Kumar Saha, Amit Kumar Das, Bhabatosh Chanda, " CBIR using Perception based Texture and Color Measures", in Proc. of 17th International Conference on Pattern Recognition(ICPR'04), Vol. 2, Aug 2004.

[2] H.B.Kekre, Sudeep D. Thepade, "Image Retrieval using Augmented Block Truncation Coding Techniques", ACM International Conference on Advances in Computing, Communication and Control (ICAC3-2009), pp. 384-390, 23-24 Jan 2009, Fr. Conceicao Rodrigous COE, Mumbai. Is uploaded on online ACM portal.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[3] H.B.Kekre, Sudeep D. Thepade, "Appraise of SPIT Problem" , SPIT-IEEE Colloquium and International Conference, 04-05 Feb 2008 , Sardar Patel Institute of Technology, Andheri, Mumbai
[4] Wang Xiaoling, " A Novel Cicular Ring Histogram for Content-based Image Retrieval", First International Workshop on Education Technology and Computer Science, 2009.
[5] Xiaoyi Song, Yongjie Li, Wufan Chen,"A Textural Feature Based Image Retrieval Algorithm", in Proc. of 4th Int. Conf. on Natural Computation, Oct. 2008.
[6] Jing Zhang, Gui-li Li, Seok-wum He, "Texture-Based Image Retrieval By Edge Detection Matching GLCM ", in Proc. of 10th International conference on High Performance Computing and Comm., Sept. 2008.
[7] R. M. Gray, "Vector quantization", IEEE ASSP Mag., pp.: 4-29, Apr. 1984.
[8] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," IEEE Trans.Commun., vol. COM-28, no. 1, pp.: 84-95, 1980.
[9] H.B.Kekre, Tanuja K. Sarode, "New Fast Improved Clustering Algorithm for Codebook Generation for Vector Quantization", International Conference on Engineering Technologies and Applications in Engineering, Technology and Sciences, Computer Science Department, Saurashtra University, Rajkot, Gujarat. (India), Amoghsiddhi Education Society, Sangli, Maharashtra (India) , 13th – 14th January 2008.
[10] H. B. Kekre, Tanuja K. Sarode, "New Fast Improved Codebook Generation Algorithm for Color Images using Vector Quantization," International Journal of Engineering and Technology, vol.1, No.1, pp.: 67-77, September 2008.
[11] H. B. Kekre, Tanuja K. Sarode, "Fast Codebook Generation Algorithm for Color Images using Vector Quantization," International Journal of Computer Science and Information Technology, Vol. 1, No. 1, pp.: 7-12, Jan 2009.
[12] H. B. Kekre, Tanuja K. Sarode, "An Efficient Fast Algorithm to Generate Codebook for Vector Quantization," First International Conference on Emerging Trends in Engineering and Technology, ICETET-2008, held at Raisoni College of Engineering, Nagpur, India, pp.: 62- 67, 16-18 July 2008. Avaliable at IEEE Xplore.
[13] H. B. Kekre, Tanuja K. Sarode, "Fast Codebook Generation Algorithm for Color Images using Vector Quantization," International Journal of Computer Science and Information Technology, Vol. 1, No. 1, pp.: 7-12, Jan 2009.
[14] H. B. Kekre, Tanuja K. Sarode, "Fast Codevector Search Algorithm for 3-D Vector Quantized Codebook", WASET International Journal of cal Computer Information Science and Engineering (IJCISE), Volume 2, No. 4, pp.: 235-239, Fall 2008. Available: http://www.waset.org/ijcise.
[15] H. B. Kekre, Tanuja K. Sarode, "Fast Codebook Search Algorithm for Vector Quantization using Sorting Technique", ACM International Conference on Advances in Computing, Communication and Control (ICAC3-2009), pp: 317-325, 23-24 Jan 2009, Fr. Conceicao Rodrigous College of Engg., Mumbai. Available on ACM portal.
[16] Jim Z.C. Lai, Yi-Ching Liaw, and Julie Liu, "A fast VQ codebook generation algorithm using codeword displacement", Pattern Recogn. vol. 41, no. 1, pp.: 315–319, 2008.
[17] H.B.Kekre, Sudeep D. Thepade, "Panoramic View Construction using Partial Images", IEEE sponsored International Conference on Sensors, Signal Processing, Communication, Control and Instrumentation (SSPCCIN-2008) 03-05 Jan 2008,Vishwakarma Institute of Technology, Pune.
[18] H.B.Kekre, Sudeep D. Thepade, "Vista creation using Picture Parts", International Conference on Emerging Technologies and Applications in Engineering, Technology and Sciences (ICETAETS-2008), 12-13 Jan 2008, Held at Computer Science Dept., Saurashtra University, Rajkot, Gujarat. (India).
[19] H. B. Kekre, Tanuja K. Sarode, Bhakti Raul, "Color Image Segmentation using Kekre's Fast Codebook Generation Algorithm Based on Energy Ordering Concept", ACM International Conference on Advances in Computing, Communication and Control (ICAC3-2009), pp.: 357-362, 23-24 Jan 2009, Fr. Conceicao Rodrigous College of Engg., Mumbai. Available on ACM portal.
[20] H. B. Kekre, Tanuja K. Sarode, Bhakti Raul, "Color Image Segmentation using Kekre's Algorithm for Vector Quantization", International Journal of Computer Science (IJCS), Vol. 3, No. 4, pp.: 287-292, Fall 2008. Available: http://www.waset.org/ijcs.
[21] H. B. Kekre, Tanuja K. Sarode, Bhakti Raul, "Color Image Segmentation using Vector Quantization Techniques Based on Energy Ordering Concept" International Journal of Computing Science and Communication Technologies (IJCSCT) Volume 1, Issue 2, pp: 164-171, January 2009.
[22] H. B. Kekre, Tanuja K. Sarode, Bhakti Raul, "Color Image Segmentation Using Vector Quantization Techniques", Advances in Engineering Science Sect. C (3), pp.: 35-42, July-September 2008
[23] H. B. Kekre, Tanuja K. Sarode, "Speech Data Compression using Vector Quantization", WASET International Journal of Computer and Information Science and Engineering (IJCISE), vol. 2, No. 4, pp.: 251-254, Fall 2008. available: http://www.waset.org/ijcise.
[24] H. B. Kekre, Ms. Tanuja K. Sarode, Sudeep D. Thepade, "Image Retrieval using Color-Texture Features from DCT on VQ Codevectors obtained by Kekre's Fast Codebook Generation", ICGST-International Journal on Graphics, Vision and Image Processing (GVIP), Volume

9, Issue 5, pp.: 1-8, September 2009. Available online at http://www.icgst.com/gvip/Volume9/Issue5/P1150921752.html

[25] H. B. Kekre, Kamal Shah, Tanuja K. Sarode, Sudeep D. Thepade, "Performance Comparison of Vector Quantization Technique – KFCG with LBG, Existing Transforms and PCA for Face Recognition", International Journal of Information Retrieval (IJIR), Vol. 02, Issue 1, pp.: 64-71, 2009.

[26] H.B.Kekre, Tanuja Sarode, Sudeep D. Thepade, "Color-Texture Feature based Image Retrieval using DCT applied on Kekre's Median Codebook", International Journal on Imaging (IJI), Volume 2, Number A09, Autumn 2009,pp. 55-65. Available online at www.ceser.res.in/iji.html (ISSN: 0974-0627).

[27] H.B.Kekre, Sudeep D. Thepade, "Image Retrieval using Non-Involutional Orthogonal Kekre's Transform", International Journal of Multidisciplinary Research and Advances in Engineering (IJMRAE), Ascent Publication House, 2009, Volume 1, No.I, 2009. Abstract available online at www.ascent-journals.com

[28] H.B.Kekre, Sudeep D. Thepade, "Color Based Image Retrieval using Amendment Block Truncation Coding with YCbCr Color Space", International Journal on Imaging (IJI), Volume 2, Number A09, Autumn 2009, pp. 2-14. Available online at www.ceser.res.in/iji.html (ISSN: 0974-0627).

[29] H.B.Kekre, Sudeep D. Thepade, "Image Retrieval using Color-Texture Features Extracted from Walshlet Pyramid", ICGST International Journal on Graphics, Vision and Image Processing (GVIP), Volume 10, Issue I, Feb.2010, pp.9-18, Available online www.icgst.com/gvip/Volume10/Issue1/P1150938876.html

[30] H.B.Kekre, Sudeep D. Thepade, "Using YUV Color Space to Hoist the Performance of Block Truncation Coding for Image Retrieval", In Proc. of IEEE International Advanced Computing Conference 2009 (IACC'09), Thapar University, Patiala, INDIA, 6-7 March 2009.

[31] H.B.Kekre, Sudeep D. Thepade, "Rendering Futuristic Image Retrieval System", In Proc. of National Conference on Enhancements in Computer, Communication and Information Technology, EC2IT-2009, 20-21 Mar 2009, K.J. Somaiya College of Engineering, Vidyavihar, Mumbai-77.

[32] H.B.Kekre, Sudeep D. Thepade, "Creating the Color Panoramic View using Medley of Grayscale and Color Partial Images ", WASET International Journal of Electrical, Computer and System Engineering (IJECSE), Volume 2, No. 3, Summer 2008. Available online at www.waset.org/ijecse/v2/v2-3-26.pdf

[33] H.B.Kekre, Sudeep D. Thepade, "Rotation Invariant Fusion of Partial Images in Vista Creation", WASET International Journal of Electrical, Computer and System Engineering (IJECSE), Volume 2, No. 2, Spring 2008. Available online at www.waset.org/ijecse/v2/v2-2-13.pdf

[34] H.B.Kekre, Sudeep D. Thepade, "Scaling Invariant Fusion of Image Pieces in Panorama Making and Novel Image Blending Technique", International Journal on Imaging (IJI), Autumn 2008, Volume 1, No. A08, Available online at www.ceser.res.in/iji.html (ISSN: 0974-0627).

[35] H.B.Kekre, Sudeep D. Thepade, "Image Blending In Vista Creation using Kekre's LUV Color Space" , SPIT - IEEE Colloquium and International Conference, 04-05 Feb 2008, SPIT Andheri, Mumbai

[36] H.B.Kekre, Sudeep D. Thepade, Archana Athawale, Anant Shah, Prathmesh Verlekar, Suraj Shirke, "Performance Evaluation of Image Retrieval using Energy Compaction and Image Tiling over DCT Row Mean and DCT Column Mean", Springer-International Conference on Contours of Computing Technology (Thinkquest-2010), Babasaheb Gawde Institute of Technology, Mumbai, 13-14 March 2010, The paper will be uploaded on online Springerlink.

[37] H.B.Kekre, Tanuja K. Sarode, Sudeep D. Thepade, Vaishali Suryavanshi,"Improved Texture Feature Based Image Retrieval using Kekre's Fast Codebook Generation Algorithm", Springer-International Conference on Contours of Computing Technology (Thinkquest-2010), Babasaheb Gawde Institute of Technology, Mumbai, 13-14 March 2010, The paper will be uploaded on online Springerlink.

[38] H.B.Kekre, Tanuja K. Sarode, Sudeep D. Thepade, "Image Retrieval by Kekre's Transform Applied on Each Row of Walsh Transformed VQ Codebook", (Invited), ACM-International Conference and Workshop on Emerging Trends in Technology (ICWET 2010),Thakur College of Engg. And Tech., Mumbai, 26-27 Feb 2010, The paper is invited at ICWET 2010. Also will be uploaded on online ACM Portal.

[39] H.B.Kekre, Sudeep D. Thepade, Archana Athawale, Anant Shah, Prathmesh Verlekar, Suraj Shirke,"Energy Compaction and Image Splitting for Image Retrieval using Kekre Transform over Row and Column Feature Vectors", International Journal of Computer Science and Network Security (IJCSNS),Volume:10, Number 1, January 2010, (ISSN: 1738-7906) Available at www.IJCSNS.org.

[40] H.B.Kekre, Sudeep D. Thepade, Archana Athawale, Anant Shah, Prathmesh Verlekar, Suraj Shirke, "Walsh Transform over Row Mean and Column Mean using Image Fragmentation and Energy Compaction for Image Retrieval", International Journal on Computer Science and Engineering (IJCSE),Volume 2S, Issue1, January 2010, (ISSN: 0975–3397). Available online at www.enggjournals.com/ijcse.

[41] http://wang.ist.psu.edu/docs/related/Image.orig (Last referred on 23 Sept 2008)

[42] S. Nene, S.Nayar, & H. Murase. Columbia object image library (COIL-100). Technical report, CUCS-006-96,

Feb-1996-http://www1.cs.columbia.edu/CAVE/software/softlib/coil-100.php.

Author Biographies

Dr. H. B. Kekre has received B.E. (Hons.) in Telecomm. Engineering. from Jabalpur University in 1958, M.Tech (Industrial Electronics) from IIT Bombay in 1960, M.S.Engg. (Electrical Engg.) from University of Ottawa in 1965 and Ph.D. (System Identification) from IIT Bombay in 1970 He has worked as Faculty of Electrical Engg. and then HOD Computer Science and Engg. at IIT Bombay. For 13 years he was working as a professor and head in the Department of Computer Engg. at Thadomal Shahani Engineering. College, Mumbai. Now he is Senior Professor at MPSTME, SVKM's NMIMS. He has guided 17 Ph.Ds, more than 100 M.E./M.Tech and several B.E./ B.Tech projects. His areas of interest are Digital Signal processing, Image Processing and Computer Networking. He has more than 270 papers in National / International Conferences and Journals to his credit. He was Senior Member of IEEE. Presently He is Fellow of IETE and Life Member of ISTE Recently seven students working under his guidance have received best paper awards. Currently 10 research scholars are pursuing Ph.D. program under his guidance.

Sudeep D. Thepade has Received B.E.(Computer) degree from North Maharashtra University with Distinction in 2003. M.E. in Computer Engineering from University of Mumbai in 2008 with Distinction, currently pursuing Ph.D. from SVKM's NMIMS, Mumbai. He has more than 06 years of experience in teaching and industry. He was Lecturer in Dept. of Information Technology at Thadomal Shahani Engineering College, Bandra(w), Mumbai for nearly 04 years. Currently working as Assistant Professor in Computer Engineering at Mukesh Patel School of Technology Management and Engineering, SVKM's NMIMS, Vile Parle(w), Mumbai, INDIA. He is member of International Association of Engineers (IAENG) and International Association of Computer Science and Information Technology (IACSIT), Singapore. His areas of interest are Image Processing and Computer Networks. He has about 69 papers in National/International Conferences/Journals to his credit with a Best Paper Award at International Conference SSPCCIN-2008 and Second Best Paper Award at ThinkQuest-2009 National Level paper presentation competition for faculty.

Tanuja K. Sarode has Received M.E.(Computer Engineering) degree from Mumbai University in 2004, currently Pursuing Ph.D. from Mukesh Patel School of Technology, Management and Engg., SVKM's NMIMS, Vile-Parle (W), Mumbai, INDIA. She has more than 10 years of experience in teaching. Currently working as Assistant Professor in Dept. of Computer Engineering at Thadomal Shahani Engineering College, Mumbai. She is life member of IETE, member of International Association of Engineers (IAENG) and International Association of Computer Science and Information Technology (IACSIT), Singapore. Her areas of interest are Image Processing, Signal Processing and Computer Graphics. She has 60 papers in National /International Conferences/journal to her credit.

Vaishali Suryawanshi has received B.E (Computer Engineering) degree from North Maharashtra University in 2000. Currently she is pursuing her M.E. (Computer Engineering) from Thadomal Shahani Engineering College Mumbai. She is working as a lecturer in Thodomal Shahani Enginnering college and has 7 years of teaching experience.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

An Efficient Trust Establishment Framework for MANETs

Mohammad Karami, Mohammad Fathian
Department of Industrial Engineering
Iran University of Science and Technology
Tehran, Iran

Abstract— **In this paper, we present a general trust establishment framework comprising three components. The first part is the trust computation model that evaluates the trust level of each participating node through monitoring and quantification of some relevant behavioral indicative metrics. The second part is the trust evidence distribution scheme that distributes the trust evidences obtained by the first component. And finally the third part is the reputation computation model that combines the collected trust evidences from other nodes to form an overall reputation score and a judgment basis regarding the trustworthiness level of each node.**

The trust computation model is based on first-hand evidences obtained via direct observations at the MAC layer. The proposed trust evidence distribution scheme is an efficient, scalable and completely distributed scheme based on ant colony optimization algorithm. For combination of collected evidences in the reputation computation model, Dempster's rule for combination is applied. Dempster's rule for combination gives a numerical procedure for fusing together multiple pieces of evidence from unreliable observers.

The paper, illustrates the applicability of the proposed framework on data packet delivery functionality with Dynamic Source Routing (DSR) as the underlying routing protocol. We present simulation results which demonstrate the effectiveness and efficiency of the proposed framework.

Keywords- Trust establishment framework; mobile ad hoc network (MANAT); evidence distribution; ant colony optimization; Dempster-Shafer theory

I. INTRODUCTION

Mobile ad hoc networks (MANETs) are multihop wireless networks spontaneously constructed by mobile nodes without relying on any pre-established infrastructure [1]. In MANETs, nodes can directly communicate with other nodes within their wireless transmission range that are often referred to as neighbors. However, to communicate with non-neighbor nodes, they have to follow a multi-hop scenario where the source nodes rely on their neighbors and several other intermediate nodes to relay their messages and deliver them to the destination. Therefore, the cooperation of participating nodes plays a vital role for successful communications. Early routing and communication protocols for MANETs have been developed optimistically, where the benign and cooperative behavior of all the participating nodes is presumed. However, it may not be always the case and in the absence of a fixed trust or security infrastructure; some nodes may decide to exhibit a non-cooperative or malicious behavior for a variety of incentives including better service, selfishness, monetary benefits or malicious intents.

Due to the unique characteristics of MANETs such as shared wireless medium, the lack of any fixed infrastructure, mobility and consequently dynamic topology changes, and resource-constrained nodes in terms of battery and computation capability, these networks are seriously susceptible to a large number of security attacks [2]. The aforementioned characteristics also prevent traditional cryptographic-based security methods to be directly applicable to MANETs.

As a result, in recent years researchers have taken a trust-based approach which promotes modeling and computing trust by defining and monitoring some behavioral indicative metrics and coming up with some sort of belief in trustworthiness level of other nodes. This computed degree of trustworthiness may then be used in situations where a node has to rely on previously unknown and therefore unreliable nodes for accomplishment of a cooperative service. In a MANET context, trust is defined as a belief level that one node can put on another node for a specific action according to previous direct or indirect information from observation of behaviors. The belief level is the extent to which one node believes that another node is willing and able to obey the protocol and act normally [3].

In this paper, we present a trust establishment framework that is based on first-hand evidences obtained via direct observations at the MAC layer as well as second-hand evidences that are obtained via an ant-based trust evidence distribution scheme from other nodes. A common difficulty in trust-based schemes that incorporate various trust evidence exchange mechanisms to reinforce their accuracy pertains to the combination of observational data from nodes that can vary in their reliability or trustworthiness. In this paper, we have employed the Dempster-Shafer evidence theory, which is well suited to an ad-hoc network where doubt and uncertainty is inherent.

The remainder of the paper is organized as follows. Section II briefly reviews related work on trust establishment in MANETs. Section III is dedicated to the details of our proposed trust establishment framework. Section IV presents results from simulation experiments that demonstrate the effectiveness of the proposed scheme. The final section of the paper discusses concluding remarks.

II. RELATED WORK

In recent years, security establishment in MANETs by the means of trust modeling and management has been a considerable topic of interest. The proposed trust management frameworks in literature fall into two major categories, reputation-based [4,5] and trust establishment [6-9]. In the former category, trust in other nodes is evaluated by direct observation and second-hand information distributed among a network. In this category most of the proposed methods use a Bayesian approach based on Beta distribution [3, 5, 10, 11]. In this approach, a random variable that follows the beta distribution is associated with the trust value of a node. Also, the posterior distribution that represents a notion of trust is derived from a prior distribution. In the later category [6-9], trust in neighbors is evaluated by direct observation, and trust relations between two nodes without previous direct interaction are established through a combination of opinions from intermediate nodes.

L. Eschenauer et al. [12] present a high-level framework for generation, revocation and distribution of trust evidence and demonstrate the significance of estimation metrics in trust establishment. A.A. Pirzada et al. [13] present a trust model that allows the evaluation of the reliability of the routes, using only first-hand information. The notion of confidence as it relates to trust management was explored by G. Theodorakopoulos et al. [14]. L. Buttyan et al. [15] propose a framework for stimulating cooperation in MANETs. The approach is based on a credit system for packet forwarding while trusted hardware is assumed.

The majority of research works presented in the literature have mainly concentrated on trust modeling and quantification, while little attention has been paid to efficient distribution of trust information. In most of the proposed trust establishment schemes participating nodes are required to periodically disseminate their trust information acquired through direct observations.

These trust information are received by other nodes and combined to form an overall reputation score for each node. This proactive approach suffers scalability, efficiency and robustness problems in resource-constrained environments [16]. Tiang and Baras [17] propose an efficient ant-based approach for the distribution of trust certificates in MANETs. However, their proposed scheme does not involve any trust or reputation computation model. In this paper we use an efficient on-demand trust evidence discovery protocol based on ant colony optimization algorithm for the distribution of trust evidences.

Yet another challenge in reputation-based schemes is related to employing an accurate, robust and straightforward method for combining observational data from nodes that can vary in their reliability or trustworthiness. Previous approaches have used simplistic combination techniques such as averaging or majority voting [18,19]. Here we apply Dempster-Shafer mathematical theory of evidence to combine independent pieces of evidence collected from other nodes in order to form an overall reputation score regarding the trustworthiness degree of a given node.

III. THE PROPOSED FRAMEWORK

As in real life, in MANETs context, trust levels are determined for particular actions. Obviously, trust computation for any action of interest requires clear definition, monitoring and quantification of some relevant behavioral indicative metrics. We believe that our proposed framework is a general framework and once corresponding metrics for a given action of interest are properly defined, monitored and quantified, it may be adapted for various scenarios. However, to give a practical illustration, for the rest of the paper, we will be particularly considering the incorporation of the proposed framework into data packet delivery functionality with Dynamic Source Routing (DSR) as the underlying routing protocol [20]. In the resulted trust-aware DSR protocol, the trustworthiness degree of intermediate nodes is taken into account, so that, non-cooperative nodes could be avoided in route selection decisions. The details of the proposed trust establishment framework are discussed in subsequent subsections.

A. *Trust Computation Model*

The trust computation model is executed by each individual node. Each node operates independently and maintains its individual perspective of the trust hierarchy. Each node uses a direct observation mechanism for monitoring data packet forwarding behavior of its neighbor nodes and accordingly quantifies trust level of each neighbor node.

In the proposed scheme, each node buffers all the packets it has sent, puts itself in promiscuous mode, initiates a timer and then overhears its neighbor's forwarding behavior. If a packet is properly forwarded within the expected timeout, then a successful forwarding event is recorded, otherwise an unsuccessful forwarding event is recorded. The trust level is simply computed by dividing the number of successful forwarding observations for a particular node by the total number of packets sent to that node to be forwarded. In particular, the trust value, *t*, assigned to node *j* by node *i* is defined as follows:

$$t_{ij} = \frac{N_s}{N_s + N_u} \tag{1}$$

Where $0 \le t \le 1$ and N_s and N_u respectively represent the cumulative number of successful and unsuccessful forwarding events of node j recorded by node i. A trust value of 0 for a given node represents complete distrust and a value of 1

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

implies absolute trust in packet forwarding functionality of that node.

The trust value computed for each neighbor node is signed by observer's private key and therefore can't be modified by intermediate nodes. We assume that the public key of the signer is well known and authenticated, and the corresponding private key cannot be compromised. Trust evidence is a foursome tuple denoted as *TE=<provider, target, TV, time>*. *Provider* is the observer node which has computed the trust value, *target* represents the node for which this trust evidence is produced, *TV* is the trust value of target node computed by the provider and finally *time* is the last update time of the trust evidence. Trust evidences are locally stored by observer nodes.

In the proposed framework as it applies to the data packet delivery functionality of DSR protocol, whenever a node needs to choose among available paths to communicate with a given destination, it first evaluates the reliability of each available path and consequently chooses the most reliable one. Path reliability is computed as the probability that a packet won't be dropped by the nodes along the route and will be safely delivered to its destination.

To compute reputation scores, a node first employs the trust evidence discovery protocol to collect relevant trust evidences and then applies the reputation computation model to combine multiple pieces of independent trust evidences collected from other nodes. The details of these two steps are discussed in following subsections.

B. *Trust Evidence Discovery Protocol*

Although there exist some literature on trust evidence discovery in P2P networks [21,22], very little attention has been paid to exclusive study of trust evidence discovery/distribution problem in MANETs. Typical approaches for trust evidence discovery in P2P networks rely on either flooding or centralized storage. The flooding approach imposes efficiency and scalability problems and the centralized storage approach is against the decentralized and infrastructure-less nature of MANETs and also imposes robustness risks.

Almost all of the trust establishment schemes that utilize trust information sharing mechanisms take a proactive approach, where nodes periodically broadcast their first-hand trust information to their neighbors. This approach also suffers scalability, efficiency and uneven distribution of trust evidences across the network.

Here we introduce an efficient on-demand ant-based trust evidence discovery protocol. Our ant-based scheme uses the swarm intelligence paradigm [23]. The swarm intelligence paradigm is inspired from artificial ant colonies techniques to solve combinatorial optimization problems [24]. The main principle behind the interaction in a swarm is called stigmergy – indirect communication through the environment. An example of stigmergy is pheromone laying on the trails followed by ants. Ants are attracted to pheromones and thereby they tend to follow the trails that have high pheromone concentrations.

The idea of the proposed ant-based scheme is inspired by the process used by real ant colony. The ant can seek path between the nest (source node) and multiple food sources (nodes hosting relevant trust evidences). They accomplish the mission with great efficiency. As the environment changes, ants can also quickly discover new routes. Since trust evidence discovery is a process to find relevant evidences with the best efficiency, utilizing the ant colony optimization proves to be helpful.

To obtain desired trust evidences hosted by other nodes, a node generates several artificial ants. The probabilistic movement of the ant allows it to explore new paths and find the proper trust evidence provider. During the trust evidence discovery period, Forward ants (***Fa***) and backward ants (***Ba***) are used. ***Fa*** is generated by trust evidence requester to explore a path to a proper trust evidence provider. ***Ba*** which contains a relevant piece of trust evidence is generated from the trust evidence provider and routes back to the requester.

The formats of ***Fa*** and ***Ba*** packets are shown in Fig. 1. The ***Fa*** packet contains *RID* – requester's ID, *TID* – target's ID (the node for which we are interested to obtain trust evidences), *SeqN* – the unique sequence number, *TTL* – the maximum number of intermediate nodes allowed to forward the ***Fa*** packet and *pass list*– the dynamically increasing list which consists of the passed nodes' IDs. In the ***Ba*** packet *PID* is the ID of trust provider node which creates the backward ant and *TimeStamp* is the creation time of the ***Ba*** packet.

RID	TID	SeqN	TTL	Pass List …

(a)

RID	PID	TimeStamp	Pass List …

(b)

Figure 1. (a) FA packet (b) BA packet

Along the path of delivering requested trust evidences, backward ants modify the information stored in the trust evidence table (TET) of each node. The structure of trust evidence table (TET) is shown in Fig. 2.

	N_1	N_2	…	N_m
TE_1	P_{11}	P_{12}	…	P_{1m}
TE_2	P_{21}	P_{22}	…	P_{2m}
…	…	…	…	…
TE_n	P_{n1}	P_{n2}	…	p_{nm}

Figure 2. Trust Evidence Table (TET)

Each row in TET corresponds to trust evidence of a node. For each trust evidence TE_n and for each neighbor node i, the probability value p_{ni} expresses the probability of choosing node i as the next hop when searching for trust evidence n and is calculated by the formula (2):

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

$$P_{ni} = \begin{cases} \dfrac{p_{ni}}{\sum_{j \in N} p_{nj}} & \text{if } i \in N \\ 0 & \text{otherwise} \end{cases} \qquad (2)$$

In (2), N is the neighbor node set of current node and p_{ni} is the amount of pheromone on the link between current node and node i for trust evidence n. During the trust evidence discovery process, p_{ni} is updated using the following formula:

$$p_{ni} = (1-\alpha).p_{ni} + \Delta p_{nj} \qquad (3)$$

Where $0 < \alpha < 1$ is the pheromone evaporation parameter, Δp_{ni} is the increment amount of p_{ni} and is determined by information contained in the received ***Ba*** and is calculated using the following formula:

$$\Delta p_{ni} = r^{-m} h^{-n} \qquad (4)$$

In the above formula, r is the recency of the trust evidence contained in the received ***Ba***, h is the hop count the ant have passed by from its source to the current node. m and n are parameters which determine the relative importance of trust evidence recency versus hop count.

To improve the performance of the trust evidence discovery protocol, discovered evidences are cached in trust evidence repository (TER) of every node on the path of backward ants. Therefore after a period of adaptation, the request overhead will be drastically reduced, since probability of obtaining required evidences from neighbors would increase. The replication procedure assures the availability of trust evidences, even when some origins may be out of reach. Upon receiving fresher trust evidences, cached evidences are updated. The cached trust evidence $TE_{i,j}$ provided by node i about node j will be deleted from the TER of the current node if a more recent evidence is not received from node i about node j in a fixed time interval Δt .

In the proposed trust establishment framework as it applies to the data packet delivery functionality of DSR protocol, relevant evidences are collected by the requester nodes in two following modes:

Implicit mode: In this mode, the trust evidence discovery process is incorporated into the route discovery mechanism of DSR protocol. Here, in addition to standard fields, each Route Reply message contains a field (trust evidence record) specially considered for recording relevant trust evidences. Before receiving a Route Request message at the target node, the route discovery process is performed according to the standard specifications of DSR protocol. When the target node received the Route Request message, it checks the list of all intermediate nodes contained in the route record of the received message, searches its TER, extract evidences related to those nodes and in addition to the standard route record, it appends these evidences to the trust evidence record of the Route Reply message that it creates. The target node then sends back the Route Reply message to the initiator node on the reverse path. Each intermediate node that forwards the Route Reply message also checks the list of nodes contained in the route record (excluding the source and destination nodes) and appends relevant trust evidences by referring to its own TER. Intermediate nodes avoid appending repetitive trust evidences and also replace recorded evidences if they have more recent versions of those evidences in their TER. Also a node forwarding the Route Reply message adds useful trust evidences to its own TER.

Explicit mode: In situations where a source node has multiple routes to a given target of communication, but due to the lack of adequate trust evidences for the nodes along the routes cannot effectively evaluate the reliability of available routes, it follows the following procedure:

1) The source node creates a forward ant Fa and broadcasts it to its neighbors.
2) Each neighbor node receiving the ***Fa*** searches its first hand trust evidence storage. If a relevant piece of evidence is found, a backward ant ***Ba*** containing the discovered trust evidence will be generated and will retrace the path of the Fa back to the source. As the ***Ba*** moves on its path, the intermediate nodes will update their TET using the formula (4) and will store the evidence in their TER.
3) After decreasing the TTL value of the received Fa, if it is still greater than zero, then the current node will unicast the Fa to the neighbor with the highest probability by consulting its TET. If there is no preference to the neighbors, i.e. there is no entry in the TET for this evidence, the Fa will be broadcasted to all neighbors. This happens either when no path to the requested trust evidence has been explored or the information of the node is outdated. Nodes discard repetitive Fa packets by checking the sequence number of received packets.
4) The requester node waits for a predefined period of time in order to get relevant trust evidences from other nodes. Once the requested evidences are received, the requester applies the reputation computation model to combine evidences related to each node to form an overall reputation score for each node.

The next section discusses the details of the reputation computation model.

C. Reputation Computation Model

In the proposed framework, nodes utilize the trust evidence discovery protocol discussed in previous section to obtain relevant trust evidences from other nodes. The obtained trust evidences are combined to form an overall reputation score for each node. Combination of trust evidences from other nodes that can vary in their reliability or trustworthiness is a challenging task and has a significant impact on the overall effectiveness of trust establishment framework. Previous approaches have used simplistic combination techniques such as averaging or majority voting [18,19]. Here, we employ the Dempster-Shafer evidence theory which offers an alternative to

 http://sites.google.com/site/ijcsis/
ISSN 1947-5500

traditional probabilistic theory for the mathematical representation of uncertainty and is well suited to our context where doubt and uncertainty are inherent. The theory and its applicability to reputation computation are discussed in subsequent subsections.

1) Dempster-Shafer Theory of Evidence

Dmpster-Shafer Theory (DST) is a mathematical theory of evidence. The seminal work on the subject is [25], which is an expansion of [26]. The theory's practical appeal is largely due to Dempster's rule for combining beliefs based on independent pieces of evidence. In a finite discrete space, Dempster-Shafer theory can be interpreted as a generalization of probability theory where probabilities are assigned to sets as opposed to mutually exclusive singletons. Let X be the universal set: the set of all states under consideration. The power set, $P(X)$, is the set of all possible sub-sets of X, including the empty set. Any hypothesis *A* will refer to a subset of power set for which observers can present evidence.

There are three important functions in Dempster-Shafer theory: the basic probability assignment function (*bpa* or *m*), the Belief function (*Bel*), and the Plausibility function (*Pl*). The *bpa*, represented by *m*, defines a mapping of each subset of the power set to the interval between 0 and 1. Formally, $m : P(X) \rightarrow [0,1]$ where it verifies two axioms. First, the mass of the empty set is zero:

$$m(\varnothing) = 0 \tag{5}$$

Second, the summation of the *bpa*s of all the subsets of the power set is 1:

$$\sum_{A \in P(X)} m(A) = 1 \tag{6}$$

The value of the *bpa* for a given set *A* (represented as *m*(*A*)), expresses the proportion of all relevant and available evidence that supports the claim that a particular element of X (the universal set) belongs to the set *A* but to no particular subset of *A*. From the basic probability assignment, the upper and lower bounds of an interval can be defined. This interval contains the precise probability of a set of interest (in the classical sense) and is bounded by two nonadditive continuous measures called *Belief* and *Plausibility*. The *Belief function (Bel)* maps a hypothesis *A* to a value between 0 and 1and is defined as follows.

$$Bel(A) = \sum_{B | B \subseteq A} m(B) \tag{7}$$

The belief function constitutes the lower bound of the interval and represents the weight of evidence supporting *A*'s provability. The *plausibility function* maps each hypothesis *A* to a value *pls*(*A*) between 0 and 1and is defined as follows.

$$pl(A) = \sum_{B | B \cap A \neq \varnothing} m(B) \tag{8}$$

The plausibility function constitutes the upper bound of the interval and represents the weight of evidence that doesn't refute *A*.

2) Dempster's Rule for Combination

Suppose $m_1(A)$ and $m_2(A)$ are the basic probability assignments from two independent observers (in the same frame of discernment). The combination (called the joint m_{12}) is calculated from the aggregation of two *bpa's* m_1 and m_2 in the following manner:

$$m_{12}(A) = \frac{\sum_{B \cap C = A} m_1(B) m_2(C)}{1-k} \tag{9}$$

Where

$$K = \sum_{B \cap C = \varnothing} m_1(B) m_2(C) \tag{10}$$

The denominator in Dempster's rule is a normalization factor and represents the basic probability mass associated with conflict.

3) Dempster's Rule for Combination Applied to Reputation Computation

We apply the Dempster's rule to combine multiple pieces of independent trust evidences collected from other nodes. In our context, the power set has three focal elements: hypothesis $H = \{T\}$ that characterizes the trust degree of a given node, hypothesis $\bar{H} = \{\bar{T}\}$ that characterizes the distrust degree of a given node and universe hypothesis $U = \{T, \bar{T}\}$ that characterizes the degree of belief that a given node is either trusted or distrusted.

For a simple illustration of how trust evidences are combined using Dempster's rule, consider that nodes A and B are offering trust evidences on node S. Assume that node A claims that trust and distrust values for S are 0.8 and 0.2 respectively and B claims that these values are 0.2 and 0.8 respectively (according to its own observations or maliciously). These two pieces of trust evidence are formalized as follows:

$$\begin{aligned} &m_A(T) = 0.8 \\ &m_A(\bar{T}) = 0.2 \\ &m_A(U) = 1 - (m_A(T) + m_A(\bar{T})) = 0 \end{aligned} \tag{11}$$

$$\begin{aligned} &m_B(T) = 0.2 \\ &m_B(\bar{T}) = 0.8 \\ &m_B(U) = 1 - (m_B(T) + m_B(\bar{T})) = 0 \end{aligned} \tag{12}$$

And the combination is computed as follows:

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

$$m_A(T)\oplus m_B(T)=\frac{m_A(T)m_B(T)+m_A(T)m_B(U)+m_A(U)m_B(T)}{1-\left(m_A(T)m_B(\bar{T})+m_A(\bar{T})m_B(T)\right)}=\frac{0.16}{0.32}=0.5 \qquad (13)$$

Dempster's rule for combination is a commutative and associative rule and therefore for any arbitrary number of *bpa*s we can compute the combination by first combining any pair of *bpa*s and then combining the result with the remaining *bpa*s in the same way.

Even though in this paper we assume that all nodes are completely reliable with respect to offering accurate trust evidences, a significant advantage of utilizing Dempster's rule for combination is its ability to effectively discount the impact of evidences obtained from unreliable sources in the computed reputation score.

An honesty coefficient with a value between 0 and 1 for the collected evidences can be utilized for this purpose. A value of 0 for a given node represents its complete dishonesty and completely neutralizes the impact of the trust evidence provided by that node in the combination rule. Conversely, a value of 1 for a given node represents its absolute honesty and maximizes the impact of the trust evidence provided by that node in the combination rule. To exemplify this, suppose that in the previous example, instead of absolute honesty, the honesty coefficient of node B was 0.8. So, we would have:

$$\begin{aligned} m_B(T) &= 0.8\times 0.2 = 0.16 \\ m_B(\bar{T}) &= 08.\times 0.8 = 0.64 \\ m_B(U) &= 1-(m_2(T)+m_2(\bar{T})) = 0.2 \end{aligned} \qquad (14)$$

And the combination rule would yield:

$$\frac{(0.8\times 0.16)+(0.8\times 0.2)}{1-\left((0.8\times 0.64)+(0.16\times 0.2)\right)}=\frac{0.288}{0.465}=0.63 \qquad (15)$$

As it can be easily verified, the impact of *B*'s trust evidence in the combination rule has been weakened and *A*'s evidence has been more influential in the gained result.

IV. Simulation and Evaluation

The Performance of the proposed framework has been evaluated using some simulations. The simulation model and gained results are discussed in following subsections.

A. *Simulation Model*

To evaluate the effectiveness and efficiency of the proposed framework, we have conducted some simulations according to multiple scenarios. We have used NS-2 for simulation purpose. All simulations are in an ad hoc network consisting of 50 nodes spread uniformly through a 1000×1000 meter square area. Nodes are equipped with an IEEE 802.11 radio network interface, operating at 11Mbps data rate with a 250m transmission range. There exist a total of 30 CBR connections and sending nodes send data packets of size 512b at 4pk/s rate. Nodes move according to the Random Waypoint mobility model with speed uniformly distributed between 0 and 10m/s and a pause time of 50 seconds. Simulations run for 900 seconds.

B. *Simulation Results*

We use the following metrics for evaluation of the proposed trust establishment framework:

Success rate: the percentage of requests for which the requester successfully obtains the relevant evidence(s). In simulation, it is the number of unique forward ants sent by the requester nodes over the total number of corresponding backward ants received by those nodes.

Throughput: In our context, throughput is defined as the ratio of the number of packets received by the application layer of destination nodes to the number of packets sent by the application layer of source nodes.

Average Latency: this metric reflects the overhead imposed by the proposed scheme. The metric is defined as the mean time (in seconds) taken by the packets to reach their respective destinations.

Fig. 3 presents the success rate performance results. As it can be observed, except at the beginning of the simulation that still trust evidences are not available and entries in trust evidence tables of participating nodes aren't accurately updated, the success rate for requested trust evidences is low. However as the simulation proceeds a fast convergence is achieved at the cost of using broadcast requests for finding desirable trust evidences.

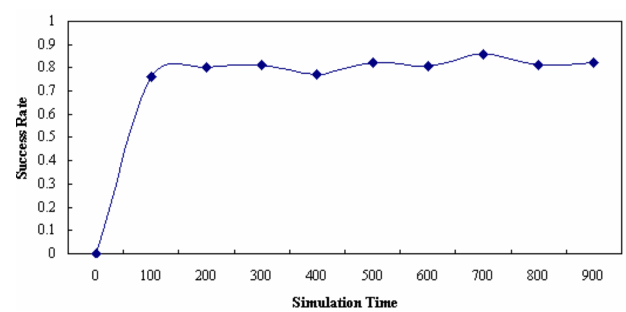


Figure 3. Success rate performance results

Notice that in fig. 3 the success rate of the proposed scheme increases and decreases repeatedly by a small amount. The reason is the mobility of nodes hosting requested trust evidences and abolishment of trust evidences with the passage of time.

To investigate the effectiveness of the proposed framework as it applies to the data packet delivery functionality of DSR protocol; we use the throughput metric. For this purpose, a varying number of selfish nodes that drop their received data packets destined to other nodes with a probability between 60 and 100 percent are implemented in simulations.

Simulations have been conducted for two different scenarios. In each scenario the ratio of selfish nodes ranges from 0 to 50 percent. The two scenarios are: 1) when standard

DSR protocol is used for route selection decisions and communications among nodes. 2) When the proposed scheme is employed to improve route selection decisions.

Fig. 4 presents the throughput performance results of the two scenarios.

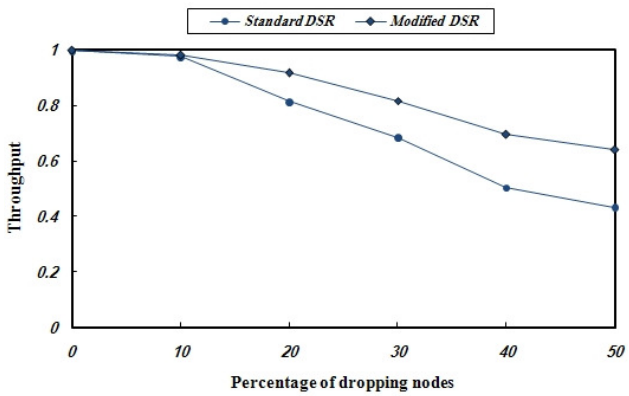


Figure 4. Throughput performance results

The performance results reveal that in the presence of a varying percentage of selfish nodes, the proposed scheme results in a better throughput. The main reason is that in standard DSR protocol, by default shorter paths are preferred for communication. However, in the proposed scheme, whenever a node needs to communicate with another node, it first evaluates the reliability of each available path and consequently chooses the most reliable one. Path reliability is defined as the probability that a packet won't be dropped by the nodes along the route and will be safely delivered to its destination.

The comparison results for average latency of standard DSR and the proposed scheme are shown in Fig 5.

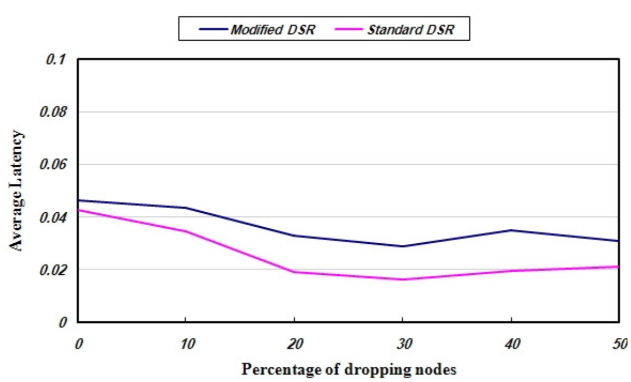


Figure 5. Average latency results

As expected, considering the fact that the trusted paths chosen by the proposed scheme are not necessarily optimal in terms of the number of hops, compared to the standard DSR protocol the average latency has been increased. However, the achieved amelioration of throughput compensates for the imposed overhead.

V. CONCLUSION

We presented a trust establishment framework that utilizes both first and second-hand observational data. An on-demand ant-based scheme was introduced for efficient distribution of trust evidences. The flexibility of the formula used for choosing the best next hop for obtaining the requested trust evidences provides the possibility of embedding more complicated metrics such as node mobility, provider's trustworthiness and security related items.

Dempster's rule for combination which offers an effective and robust mechanism for combination of trust evidences collected from other nodes and quantification of reputation was used. We used experimental simulations to demonstrate the effectiveness of the proposed framework as it applies to the data packet delivery functionality of DSR protocol. However, we believe that the proposed framework is a general framework that may be adapted for a variety of scenarios where nodes have to rely on unreliable nodes to accomplish a cooperative service.

ACKNOWLEDGMENT

This research has been supported by Iran Telecommunication Research Center (ITRC) under contract 19230/500 and herein authors are willing to express their gratitude.

REFERENCES

[1] Chlamtac, I., Conti, M. and Liu, J.N. (2003). Mobile ad hoc networking: imperatives and challenges. Ad Hoc Networks, 1, 13-64.

[2] Wu, B., Chen, J., Wu, J. and Cardei, M. (2007). A Survey of Attacks and Countermeasures in Mobile Ad Hoc Networks. In: Xiao, Shen and Du (Eds), *Wireless Network Security*, 103-136.

[3] Li, J., Li, R. and Kato, J. (2008). Future Trust Management Framework. *Communications Magazine IEEE*, 46(4), 108-114.

[4] Buchegger, S. and Le Boudec, J. Y. (2002). Performance Analysis of the CONFIDANT Protocol (Cooperation Of Nodes Fairness in Dynamic Ad-hoc NeTworks). Proc. ACM MobiHoc 2002, Atlanta, GA.

[5] Buchegger, S. and Le Boudec, J. Y. (2004). A Robust Reputation System for P2P and Mobile Ad-Hoc Networks. Proc. P2PEcon 2004, Harvard Univ., Cambridge, MA.

[6] S. Marti et al., "Mitigating Routing Misbehavior in Mobile Ad Hoc Networks," Proc. MobiCom 2000, Aug.2000, pp. 255–65.

[7] Sun, Y., Yang, Y. (2006). A Trust Evaluation Framework in Distributed Networks: Vulnerability Analysis and Defense Against Attacks. Proc. IEEE INFOCOM 2006, Barcelona, Spain.

[8] Theodorakopoulos, G. and Baras, S. (2006). On Trust Models and Trust Evaluation Metrics for Ad Hoc Networks. *Selected Areas in Communications, IEEE*, 24(2), 318-328.

[9] Zouridaki, C., Mark, B.L., Hejmo, M., Thomas, R.K. (2007). Hermes: a quantitative trust establishment framework for reliable data packet delivery in MANETs. *Journal of Computer Security*, 15 (1), 3–38.

[10] Ganeriwal, S. and Srivastava, M. (2004). Reputation-based Framework for High Integrity Sensor Networks. Proc. ACM Wksp. Sec. Ad Hoc and Sensor Networks, Washington, DC.

[11] Zouridaki, C., Mark, B.L., Hejmo, M. and Thomas, R.K. (2005). A Quantitative Trust establishment Framework for Reliable Data Packet Delivery in MANETs. Proc. 3rd ACM Wksp. Sec. Ad Hoc and Sensor Networks, 1-10.

[12] Eschenauer, L., Gligor, V.D., Baras, J. (2002). On trust establishment in mobile ad-hoc networks. proc. Wksp. Security, vol. 2845, LNCS, 47–66.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[13] Pirzada, A.A., McDonald, C. (2006). Establishing trust in pure ad-hoc networks. *Wireless Personal Communications*, 37, 139-163.

[14] Theodorakopoulos, G., Baras, J.S. (2004). Trust evaluation in ad-hoc networks. proc. ACM Wksp. Wireless Security (WiSe'04), 1–10.

[15] Buttyan, L., Hubaux, J.P. (2003). Stimulating cooperation in self-organizing mobile ad hoc networks. *Mobile Networks and Applications*, 8 (5), 579–592.

[16] Huo, H., Gao, D., Niu, Y. and Gao, S. (2007). ASDP: An Action-Based Service Discovery Protocol Using Ant Colony Algorithm in Wireless Sensor Networks, Lecture Notes in Computer Science 4864, 338-349.

[17] Jiang, T. and Baras, J.S. (2004). Ant-based Adaptive Trust Evidence Distribution in MANET. Proc. 24th International Conference on Distributed Computing Systems Workshops (ICDCSW'04), 588-593.

[18] Kargl, F., Klenk, A., Weber, M. and Schlott, S. (2004). Sensors for Detection of Misbehaving Nodes in MANETs. Proc. Detection of Intrusion and Malware and Vulnerability Assessment.

[19] Zhang, Y. and Lee, W. (2000). Intrusion Detection in Wireless Ad-Hoc Networks. Proc. 6th Ann. ACM Int'l Conf. Mobile Computing and Networking, ACM Press, 275–283.

[20] Johnson, D.B., Maltz, D.A. and Hu, Y. (2003). The Dynamic Source Routing Protocol for Mobile Ad-Hoc Networks (DSR), IETF MANET, Internet Draft (Work in Progress).

[21] Clarke, I., Sandberg, O., Wiley, B. and Hong, T.W. (2000). Freenet: A distributed Anonymous Information Storage and Retrieval System. In Proc. ICSI Workshop on Design Issues in Anonymity and Unobservability, Berkeley, CA.

[22] Babaoglu, O., Meling, H. and Montresor, A. (2002). Anthill: A Framework for the Development of Agent-Based Peer-to-Peer Systems. In Proc. 22nd ICDCS, Vienna, Austria.

[23] Bonabeau, E., Dorigo, M. and Theraulaz, G. (1999). Swarm Intelligence – From Natural to Artificial Systems. New York: Oxford University Press.

[24] Caro, G. D. and Dorigo, M. (1998). AntNet: Distributed Stigmergetic Control for Communications Networks, *Journal of Artificial Intelligence Research*, 9, 317-365.

[25] G. Shafer, A. (1976). Mathematical Theory of Evidence. Princeton: Princeton Univ. Press.

[26] Dempster, A. (1967). Upper and Lower Probabilities Induced by a Multivalued Mapping. *Ann.Mathematical Statistics*, 38(2), 325–339.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Fault Analysis Attacks and Its Countermeasure using Elliptic Curve Cryptography

M.Prabu
Research Scholar
Anna University Coimbatore
Tamil Nadu, India
+91 99422 71899
prabu_pdas@yahoo.co.in

R.Shanmugalakshmi
Assistant Professor/CSE
Government College of Technology
Tamil Nadu, India
+91 422 2432221
shanmuga_lakshmi@yahoo.co.in

***Abstract*-In the last decade, many researchers had published the overall analysis attacks of cryptographic devices against implementation on elliptic curve attacks. Usually such type of information is not sufficient to learn about the individual attacks. Now in this article, we indisputably concentrated on fault analysis attack and its countermeasure.**

Key words-components:
Elliptic Curve, Implementation Attacks, Individual attack, Fault analysis attack

I. INTRODUCTION

In research field, the embedded based devices made an enormous role on security. A lot of attention has been paid to the problem of errors occurring in cryptographic devices, such as crypto processors. The cryptographic security is mainly activated through the field of study and the study of implementation. The implementation is a major obsession to known about the overall performance clearly. When comparing the real world applications, the embedded devices use cryptographic algorithm to achieve a chief safety.

Fault analysis attacks take advantage of errors that occur while cryptographic device is performing a private- key generation. Fault analysis is one kind of side channel analysis that collects data such as time and power consumption emitted by the device during computation with private key.

II.TYPES OF FAULT

The fault types can be classified as permanent and transient

A. Permanent

In a permanent fault, it directly affects to or change ROM and code can be damaged. It is more powerful than temporary faults [1]. It is very hard to recover the permanent fault and it very hard to change or modify the damaged part.

B.Transient faults

In a transient faults [1][2], it can disturb the code of execution of a particular event. It is a tedious work to defeat the transient faults. It accesses the program counter and different execution might be executed.

III. MODEL OF FAULT ATTACKS

A. Bit versus Byte errors

The frequency to alter a value of one bit or one byte. Byte model directly affects the whole memory storage. Compare to the bit model, bit model is not induced. Because it is tedious to identify the bit level errors. [9]

B. Specific versus Random values

The frequencies of possibility to alter the value of data in specify or random but as a binary values. A random values execution is easier to induce.

C. Static versus Computational errors

Normally, the attackers can make errors in execution or computation period. After that execution, the errors are static can't able to change [8]. Computation error is easy to add in the real time process. At the similar time, it is tough to add the modified value in memory

D. Data versus Control errors

A control error occurs when some iteration are stopped because of faults. Control error is more powerful than data error. It can be very prevailing while execution period.

IV. FAULT ATTACKS AND FAULT INJECTION

In which place or in what type of situation Fault Injection has been discovered [5]

- Laser produce similar effects as multi chromatic light but allow targeting a more precise circuit area.
- Photoelectric effects due to intense light induce currents in the electric circuit [5].

To protect Fault analyses attacks, the following countermeasure could be realized in hardware or software

- Light Detectors
- Supply Voltage Detectors

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

- Frequency Detectors
- Hardware redundancy with comparison
- Checksum.

V. TYPES OF FAULT ATTACKS

A. Biehl-Major –Muller Attacks

By inserting or disturbing representation of a point on a strong elliptic curve E, and insert a random register fault on the device. Its computation to a value which is not a point on curve E but on different curve. The result of these computations is a point on the new, but not less cryptographically strong curve[3]. This can be exploited to compute the secret key d. The incorrect output values are used to compute possible intermediate values of the computation and part of the secret key.

B. Random access of the multiplication Algorithm

The cryptographically strong elliptic curve E is defined over Fq. E(Fq) contains a subgroup of prime order p with p > q/log q[7]. The multiplication operation dP =Q is done by the Binary method (Algorithm 1)

Algorithm 1(Binary method)
Input: d = (dn−1. . . d0), P€ E(Fq)
Output: dP

```
Initialize:
H = P
Q = O
for i = 0 to n − 1 do
if d_i == 1 then
Q = Q + H
end
H = 2H
return Q
end
```

Qi and Hi are the values stored in Q, H before interaction i. The following steps are using to iterate the process

step 1:

Q=Qn=dP where n=log2d

step 2:

The computation with P and enforce a fault and get a fault out put Qn.

Qn→ a bit flip in a random iteration i, such that Qj→ -Qj.[ ]

step 3:

Guessing and comparing with the know values Qn,-Qn, and Countermeasures for Multiplication Algorithm:

- Consistency of output point
- Any point which serves as basis for the computation.

C. Invalid Curve Attack

The collection of small subgroup attacks are called Invalid curve attack, which can be developed by Differential Fault Attack of Bihel, Meyer and Muller Standardized elliptic curve key establishment and public key encryption protocols such as EC-DH, DC-IES and IC-MQV [4], which are effective if the receiver of an elliptic curve point doesn't verify that the point lies on the appropriate elliptic curve.

D. Small subgroup Attack

Small Subgroup Attack Lim and Lee demonstrated in [16] the importance of public key validation by presenting small subgroup attacks on discrete-logarithm Key-agreement and encryption protocols such as Diffie-Hellman-type key exchange protocols and applications of El-Gamal encryption and signature schemes. The attacks succeed if the receiver of a group element does not verify that the element belongs to the desired group of high order. Their attacks are effective if the cofactor has many small factors [15]. The attacker can then determine the victim's secret key modulo of these small factors and combine the results using CRT.The attacks that use a faulted public key can be prevented by public key validation or by partial validation as recommended in [14] [15].

VI. SIGN CHANGE FAULTS

Sign changes of points can be used to recover the secret scalar factor d: Q = dP .The sign change curve scalar factor was identified by biomer, Otto and Seifert showed in [12]. The faulty output is a valid point on the curve and the secret scalar factor can be recovered in polynomial time. They also presented a countermeasure that is motivated by a similar countermeasure by Shamir [13].

Algorithm:
Set n: = l(k)
Set Qn: = O 2 for i from n-1 to 0 do
Set Q′i := 2 · Qi+1………………………………………. Q′i = − Q′i
If (ki = 1) then set Qi: = Q′i + P
else set Qi: = Q′i
Return Q_0
Step 1: Describe faulty final result
Qi= −Qi + 2 · Li(k),
Step 2: Collect many faulty final results
Choose block size m € O(2m) operations:
Mount (n/m) log (2n) many attacks to hit every possible block with Probability[10]. At least 1/2
Step 3: incremental computation of k
Assumption: all s lowest bits of k are known try all possibilities with up to s + m bits:
Qi !=−Qi + 2 · Ls+m−1(k)
Compare to gathered faulty final results

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

VII. COUNTERMEASURES AGAINST FAULT ATTACKS

Run the encryption twice and out put the results only if these two are identical. The main approach is that increases computation time, also the probability that fault will not occur twice is not sufficiently small. The probability of fault occurrence makes twice the functions such as encryption and decryption [2][11], this level of countermeasure is hard to implement but not impossible.

VIII.CONCLUSION

This article provides a brief explanation about the fault analysis attacks, basic performance and their countermeasures implementations. It also provides guidance for further researchers by refereeing each subtopic with more clearly.

REFERENCES

[1]. Mathieu Ciet and Marc Joye. Elliptic curve cryptosystems in the presence of permanent and transient faults. 2003. http://eprint.iacr.org/2003/028

[2] C. Clavier and M. Joye. Universal exponentiation algorithm a first step towards provable spa-resistance. In Cryptographic Hardware and Embedded Systems - CHES 2001, volume 2162 of Lecture Notes in Computer Science, page 300308. Springer-Verlag, 2001.

[3] Ingrid Biehl, Bernd Meyer, and Volker M¨uller. Differential fault attacks on elliptic curve cryptosystems. In CRYPTO '00: Proceedings of the 20th Annual International Cryptology Conference on Advances in Cryptology, pages 131–146, London, UK, 2000. Springer-Verlag

[4]. N.B. Smart. The discrete logarithm problem on elliptic curves of trace one. Journal of Cryptology, 12:193–196, 1999.

[5]. Eli Biham and Adi Shamir. Differential fault analysis of secret key cryptosystems. In CRYPTO '97: Proceedings of the 17th Annual International Cryptology Conference on Advances in Cryptology, pages 513–525, London, UK, 1997. Springer-Verlag

[6]. F. Crowe, A. Daly, and W. Marnane, "A Scalable Dual Mode Arithmetic Unit for Public Key Cryptosystems," *IEEE International Conference on Information Technology: Coding and Computing (ITCC)*, vol. 1, pp. 568 – 573, 2005.

[7]. C. Giraud. An rsa implementation resistant to fault attacks and to simple power analysis. Volume 55, pages 1116–1120, 2006

[8]. Sung-Ming Yen and Marc Joye. Checking before output may not be enough against fault based cryptanalysis. IEEE Trans. Computers, 49(9):967–970, 2000

[9]. D. Knuth and A. Yao, "Analysis of the subtractive algorithm for greatest common divisors," *Proc. Nat. Acad. Sct*, vol. 72, no. 12, pp. 4720–4722, 1987

[10]. H. Bar-El, H. Choukri, M. Tunstall, and C. Whelan. The sorcerer's apprentice guide to fault attacks. Workshop on Fault Detection and Tolerance in Cryptography - FDTC 2004, 2004

[11]. Chen Z, Zhou Y. Dual-rail random switching logic: a countermeasure to reduce side channel leakage. In: Cryptographic hardware and embedded systems – CHES 2006. Lecture notes in computer science, vol. 4249. Springer; 2006. p. 242–54

[12] Johannes Bl¨omer, Martin Otto, and Jean-Pierre Seifert. Sign change fault attacks on elliptic curve cryptosystems. Cryptology ePrint Archive, Report 2004/227, 2004. .

[13] A. Shamir. Method and apparatus for protecting public key schemes from timing and fault attacks. November 1999. US Patent No. 5,991,415..

[14] Certicom Research. Standards for Efficient Cryptography Group (SECG), SEC 1: Elliptic Curve Cryptography, September 2000. http://www.secg.org/collateral/sec1 final.pdf.

[15] ANSI X9.63. Public Key Cryptography For The Financial Services Industry: Key Agreement and Key Transport Using Elliptic Curve Cryptography, January 1999. http://grouper.ieee.org/groups/1363/private/x9- 63-01-08-99.pdf.

[16] Chae Hoon Lim and Pil Joong Lee. A key recovery attack on discrete log-based schemes using a prime order subgroup. Volume 1294, pages 249–263. Springer-Verlag, 1997

AUTHORS PROFILE

M.Prabu is working as a Lecturer in the Department of Computer Science and Engineering in Adhiyamaan college of Engineering, Hosur, Tamil Nadu, India. He has published more than 5 International/National journals.He is presently doing his Ph.D in Anna University, Coimbatore, India. His area of interest are computer Networks, Information Security and Cryptography. He is life member of ISTE.

Dr. R.Shanmugalakshmi is working as an Assistant Professor in the Department of Computer Science and Engineering in Government College of Technology, Coimbatore, India. She has published more than 40 International/National journals. Her research area includes Image Processing, Neural Networks, Information Security and Cryptography. She has received Vijya Ratna Award from India International Friendship Society in the year of 1996, she has received Mahila Jyothi Award from Integrated Council for Socio-Economic Progress in the year of 2001 and she has received Eminent Educationalist Award from International Institute of Management, New Delhi in the year of 2008.She is member of Computer Society of India, ISTE and FIE.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

A Compressed Video Steganography using Random Embedding Scheme

Sherly A P
TIFAC CORE in Cyber Security
Amrita Vishwa Vidyapeetham,
Coimbatore, India
sherlyram@gmail.com

Sapna Sasidharan
TIFAC CORE in Cyber Security
Amrita Vishwa Vidyapeetham
Coimbatore, India
sapnapv@gmail.com

Amritha P P
TIFAC CORE in Cyber Security
Amrita Vishwa Vidyapeetham
Coimbatore, India
ammuviju@gmail.com

Abstract−Steganography is the art of hiding while the communication is taking place, by hiding information in other information. Many different carrier file formats can be used, images, videos, audios, image etc. This paper proposes a Compressed Video Steganographic Scheme. In this scheme, data hiding operations are executed entirely in the compressed domain. Here data are embedded in the macro blocks of I frame with maximum scene change. To enlarge the capacity of the hidden secret information and to provide an imperceptible stego-image for human vision, random embedding scheme (Pixel Value Differencing) is used. Decompression process is not required in this scheme. Experimental results demonstrate that the proposed algorithm has high imperceptibility and capacity.

Keywords- Video Steganography; MPEG-4; PVD

I. INTRODUCTION

The quick growth of the Internet and multimedia communication systems in the past decade has enabled users to send digital data over network suitably. However, transmission of data in an open network is not secure, and data can be easily tampered by illegal users. Consequently, shielding data during transmission is an important task. Although cryptographic techniques can be used for this purpose, they are not secure enough because encryption can provide secure delivery of digital content, but when the content is decrypted, encryption no longer provides any security. To solve this problem, data hiding techniques were proposed and have been considered widely in various fields like covert communication, copyright protection, and broadcast monitoring and military communication. Steganography is the art of hiding information in such a way that no one can realize a hidden message in the data except the sender and the intended recipient. Steganography is also known as 'covered writing' which includes methods of transmitting secret messages through inoffensive cover mediums in such a manner that the survival of the embedded messages is undetectable. It can also be viewed as a tradeoff between detectability, robustness, and bit rate. Detectability is the apprehension of clandestine transmission and is often used in combination with encryption. It is robust to all types of processing such as transformations, filtering, truncation, and scaling. Finally, bit rate or the maximum amount of data that can be transmitted. This paper considers data embedding in videos. A video can be viewed as a sequence of still images and data embedding in images seems very similar to videos. However, there are many differences between data hiding in images and videos, where the first important difference is the size of the host media. Since videos contain more sample number of pixels or the number of transform domain coefficients, a video has higher capacity than a still image and more data can be embedded in the video. Also, there are some characteristics in videos

which cannot be found in images as perceptual redundancy in videos is due to their temporal features. Here data hiding operations are executed entirely in the compressed domain [1] [2]. On the other hand, as a really higher amount of data must be embedded in the case of video sequences, there is a more demanding constraint on real-time effectiveness of the system. Furthermore, with the development of multimedia and stream media on the Internet, transmitting video on the Internet will not incur suspicion. Image-based and video-based steganographic techniques are mainly classified into spatial domain and frequency domain based methods. The former embedding techniques are LSB, matrix, embedding [5] etc. Two important parameters for evaluating the performance of a steganographic system are capacity and imperceptibility. Capacity refers to the amount of data that can be hidden in the cover medium so that no perceptible distortion is introduced. Imperceptibility or transparency represents the invisibility of the hidden data in the cover media without degrading the perceptual quality by data embedding. Security is the other parameter in the steganographic systems, which refers to an unauthorized person's inability to detect hidden data. Previous work in data hiding field cared little about capacity and had low embedding capacity.

In this paper, we propose a secure compressed video steganographic architecture taking account of video statistical invisibility. This paper is organized as follows: Section II describes the framework of our video steganography system. In Section III, the embedding mechanism is described in detail. We give the experimental results in Section IV. In Section V, a conclusion is drawn finally.

II. ARCHITECTURE

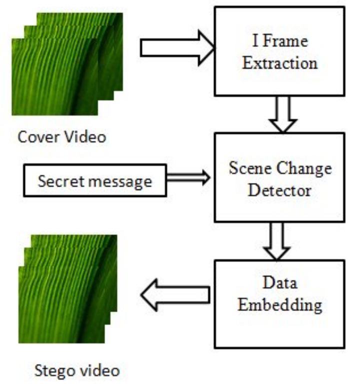


Fig.1: Block Diagram

As shown in the Fig. 1, the architecture consists of four functions: I frame extraction, the scene change detector, the data embedder. The first section explains the extraction of I frames from MPEG video. In the second module, scene change detector analyzes the frames with maximum scene change. I frames in MPEG standard is coded in intra frame manner, we can obtain the DC picture with abstracting the DC coefficients from the DCT coefficient codes. Eq.1 describes the compare method between two conjoint I frames

$$H(Ii,Ii+1)=\sum_{k=1}^{N}(Hi(k)-Hi+1(k))^\wedge 2/(Hi(k)+Hi+1(k))^\wedge 2$$

where I and, I_{i+1} means the i^{th} and $i+1^{th}$ I frames, H_i and $H_{i+1}$are histograms of DC pictures from the i^{th} and $i+1^{th}$ I frames. If the HD ($I_{i,}$ I_{i+1}) is the peak value the two I frames are from different scenes, therefore the scene change point is found. Also the variances var (i) U of each DC picture from I frame will be calculated. With the third module, data embedder, secret message M is hidden into the compressed video sequence without bringing perceptive distortion. To increase the capacity of the hidden secret information and to provide an imperceptible stego-image for human

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

vision, here pixel-value differencing (PVD) is used for embedding.

III. EMBEDDING MECHANISM

A. Embedding Position Selection

Compressed video sequence achieves compression through the elimination of temporal, spatial and statistical redundancies with the use of motion compensation, block quantization inside a discrete cosine transform (DCT), and Huffman run-level encoding. While selection of the embedding block calculate the maximum scene change of each block of the conjoint I frames. Select the block with which maximum scene change occurs by using threshold value. Selection of proper color channel is another issue in Video Steganography. The frames of video sequence is split into Y,Cb,Cr channels in the MPEG coding stage. According to different color resampling rules, it is possible for the ratio Y:Cb:Cr to be set to 4:2:2 or 4:2:0. Under these the only unchanged channel is the Y channel. So here Y channel is preferred as the host channel. In addition to the above selections, choosing inappropriate frame type among I-frame, P-frame or B-frame for hiding message is also a crucial issue. Usually, a conventional video consists of a number of GOPs. Each GOP is composed of one I-frame and several B-frames and P-frames. A typical I-frame adopts intra coding, which means it does not refer to any other frames. Different from an I-frame, a P-frame only refers to its nearest preceding I- or P-frame. As for a B-frame, it refers to the nearest preceding and succeeding I-frame or P-frame. In a conventional MPEG format, the content of a B- or P-frame is the so-called residual error between the current frame and the frame to which it refers. Therefore, only an I-frame can hold complete information. In this paper, we choose to embed message into the I-frames of an MPEG compressed video sequence.

B. Compressed Video Steganographic Algorithm

Here a novel steganographic approach called pixel-value differencing algorithm (PVD) is used for embedding. Images are more easily noticed by human eyes. In the PVD embedding method, the cover image (I frame) is simply divided into a number of non-overlapping two-pixel blocks. Each block is categorized according to the difference of the gray values of the two pixels in the block. A small difference value indicates that the block is in a smooth area and a large one indicates that it is in an edged area. The pixels in edged areas may, as mentioned previously, tolerate larger changes of pixel values than those in the smooth areas. So, in this method more data is embedded in edged areas than in the smooth areas. And it is in this way that the changes in the resulting stego-image are kept unnoticeable.

C. Data Embedding and Extraction

1) Calculate the difference value d_i between two consecutive pixels p_i and p_{i+1} for each block in the cover image. The value is given by $d_i = g_{i+1} - g_i$

2) Using d_i to locate a suitable R_k in the designed range table, that is to compute $j = \min (u_k - |d_i|)$ where $u_k >= d_i$ for all $1 <= k <= n$. Then R_j is the located range.

3) Compute the amount of secret data bits t that can be embedded in each pair of two consecutive pixels by R_j. The value t can be estimated from the width w of R_j, this can be defined by $t = log_2 w_j$

4) Read t bits from the binary secret data and transform the bit sequence into a decimal value b. For

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

instance, if bit sequence = 110, then the converted value $b = 6$.

5) Calculate the new difference value d to replace the original difference

$$d_i' = \begin{cases} l_j + b\ , & \text{if } d_i \geq 0 \\ -(l_j + b)\ , & \text{if } d_i < 0 \end{cases}$$

6) Modify the values of p_i and p_{i+1} by the following formula (g_i', g'_{i+1}) = (g_i − *ceil* (m), g_{i+1} + *floor* (m)) if d is odd

(g_i', g'_{i+1}) = (g_i − *floor* (m), g_{i+1} + *ceil* (m)) if d is even, where $m = (d' - d)/2i+1$

Repeat Step 1-6 until all secret data are embedded into the cover image, then the stego-image is obtained.

During the phase of secret extraction, the original designed range table is required. In the beginning, the same method in the embedding phase is used to partition the stego-image into pixel pairs (blocks). Then the difference value d for each pair of two consecutive pixels p_i* and p* the stego-image is calculated. Next, d_i* is used to locate the suitable R_{i+1} in Step 2 during the embedding phase. Therefore, b* is obtained by subtracting l_j from d_i*. If the stego-image is not altered, b* is equal to b. Finally, b* is transformed from a decimal value into a binary sequence with t bits, where $t = \log_2 w_j$

IV. EXPERIMENTAL RESULT

Several experiments were performed to evaluate the performance of the proposed steganographic algorithm

A. Capacity and PSNR

The secret binary data sequence S is generated by pseudo-random numbers. We set the designed range table with the width in the set of $w_k \in \{8, 8, 16, 32, 64, 128\}$. Here, PSNR value is utilized to evaluate the invisibility of the stego-images. Table I lists the experimental results after the secret data are embedded using those two approaches. The hiding capacity (in bytes) and PSNR values achieved by the proposed scheme for I-frames are shown. The listed values are the average results after embedding 100 randomly generated bit sequences into the cover images. Two stego-images are still hardly observed that the secret data is hidden inside. This is because of the high variance existed in the pixel values of the I-frame. Therefore, this demonstrates that the proposed approach provides a promising performance in increasing the capacity of the stego-images and maintaining the imperceptible quality simultaneously. This table explains the capacity and the PSNR value after embedding.

Table I
PSNR and Capacity of stego I-frames

Cover I-frames	Capacity (bytes)	PSNR (db)
I_1	70235	42.14
I_9	71345	41.5
I_{16}	70356	43.5

Fig. 4.1: I-frame before embedding

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Fig. 4.2: I-frame after embedding

V. CONCLUSION

A new Video Steganographic Scheme was proposed in this paper, operating directly in compressed domain. For data hiding pixel- value differencing (PVD) algorithm has been used. This algorithm provides high capacity and imperceptible stego-image for human vision of the hidden secret information. Here I-frame with maximum scene change blocks was used for embedding. The performance of the steganographic algorithm is studied and experimental results shows that this scheme can be applied on compressed videos with no noticeable degradation in visual quality.

REFERENCES

[1] F Hartung, B. Girod "Watermarking of Uncompressed and Compressed Video", Signal Processing, Special Issue on Copyright Protection and Access Control for Multimedia Services , 1998, 66 (3): 283-301.

[2] Bin Liu, Fenlin Liu, Chunfang Yang and Yifeng Sun, "Secure Steganography in Compressed Video Bitstreams", The Third International Conference on Availability, Reliability and Security ,2008

[3] Y. K. Lee, L. H. Chen, "High capacity image steganographic model," IEE Proceedings on Vision, Image and Signal Processing, Vol. 147, No.3, pp. 288-294, 2000.

[4] D.C. Wu, and W.H. Tsai, "A Steganographic Method for Images by Pixel-Value Differencing," Pattern Recognition Letters, Vol. 24, pp. 1613–1626, 2003

[5] Y. J. Dai, L. H. Zhang and Y. X. Yang, "A New Method of MPEG Video Watermarking Technology", International Conference on Communication Technology Proceedings (ICCT), 2003.

[6] G. C. Langelaar and R. L. Lagendijk, "Optimal Differential Energy Watermarking of DCT Encoded Images and Video", IEEE Trans. on Image Processing, 2001, 10(1):148-158.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Selective Image Encryption Using DCT with Stream Cipher

Sapna Sasidharan
TIFAC CORE in Cyber Security
Amrita Vishwa Vidyapeetham
Coimbatore, India
sapnapv@gmail.com

Jithin R
TIFAC CORE in Cyber Security
Amrita Vishwa Vidyapeetham
Coimbatore, India
jithinr550@gmail.com

Abstract—Encryption is used to securely transmit data in open networks. Each type of data has its own features; therefore different techniques should be used to protect confidential image data from unauthorized access. In this paper, selective image encryption using DCT with Stream Cipher is done. In the DCT method, the basic idea is to decompose the image into 8×8 blocks and these blocks are transformed from the spatial domain to the frequency domain by the DCT. Then, the DCT coefficients correlated to the lower frequencies of the image block are encrypted using the RC4 Stream Cipher. The resulted encrypted blocks are shuffled using the Shuffling Algorithm. Selective encryption is a recent approach where only parts of the data are encrypted to reduce the computational requirements for huge volumes of images.

Keywords- DCT; Stream Cipher; Shuffling Algorithm; Selective Encryption

I. INTRODUCTION

Currently, information security is becoming more essential in data storage and transmission. Images are widely used in several processes. Therefore, the protection of image data from unauthorized access is important. Image encryption plays a significant role in the field of information hiding. Image hiding or encrypting methods and algorithms range from simple spatial domain methods to more complicated and reliable frequency domain [1] ones. It is argued that the encryption algorithms, which have been originally developed for text data, are not suitable for securing many real-time multimedia applications because of large data sizes. A major recent trend is to minimize the computational requirements for secure multimedia distribution by "selective encryption" where only parts of the data are encrypted. Selective encryption aims at avoiding the encryption of all bits of a digital image and yet ensuring a secure encryption. The key point is to encrypt only a small part of the bit stream to obtain a fast method [2].

Several selective encryption methods have been proposed for DCT compressed images. Droogenbroeck and Benedett [3] selected AC coefficients from compressed images for encryption. In their method the DC coefficients are not ciphered because they carry important visible information and they are highly predictable. The compression and encryption stages are separated in this approach and this requires an additional operating cost. Fitch et al. [4] have proposed a partial image encryption where the data are organized in a scalable bitstream form [5]. The variety of applications for secure multimedia requires either full encryption or selective encryption. However, there is a huge spectrum of applications that

demands security on a lower level, as for example that ensured by selective encryption (SE). Such approaches reduce the computational requirements in networks with diverse client device capabilities. The goal of SE is to encrypt a well defined range of parameters or coefficients. The security level of SE is always lower when compared with the full encryption. However, SE decreases the data size to be encrypted and consequently requires lower computational time. Confidentiality is very important for lower powered systems such as for example wireless devices. Always, when considering image processing applications on such devices we should use minimal resources. However, the classical ciphers are usually too slow to be used for image and video processing in commercial low powered systems. The selective encryption (SE) can fulfill the application requirements without the overhead of the full encryption. In the case of SE, only the minimum necessary data are ciphered [6]. In [7] a technique was proposed, called zigzag permutation applicable to DCT-based videos and images. On one hand this method provides a certain level of confidentiality, while on the other hand it increases the overall bit rate. Combining SE and image/video compression using the set partitioning in hierarchical trees was used in [8]. However, this approach requires a significant computational complexity. A method that does not require significant processing time and which operates directly on the bit planes of the image was proposed in [9]. An approach that turns entropy coders into encryption ciphers using statistical models was proposed in [10]. In [11] it was suggested a technique that encrypts a selected number of AC coefficients. The DC coefficients are not ciphered since they carry important visual information and they are highly predictable. In spite of the constancy in the bit rate while preserving the bit stream compliance, this method is not scalable. Moreover, the compression and the encryption process are separated and consequently the computational complexity is increased [6].

II. DISCRETE COSINE TRANSFORM

The DCT is a mathematical transformation that takes a signal and transforms it from spatial domain into frequency domain. Many digital image and video compression schemes use a block-based DCT, because this algorithm minimizes the amount of data needed to recreate a digitized image. In particular, JPEG and MPEG use the DCT to concentrate image information by removing spatial data redundancies in two-dimensional images [12]. In the standard JPEG encoding, the representation of the colors in the image is converted from RGB to YCbCr, then the image is decomposed in 8×8 blocks, these blocks are transformed from the spatial to the frequency domain by the DCT. Then, each DCT coefficient is divided by its corresponding constant in a standard quantization table and rounded down to the nearest integer. After this step, the DCT quantized coefficients are scanned in a predefined zigzag order to be used in the final step, the lossless compression as illustrated in Fig. 1. In each block the 64 DCT coefficients are set up from the lowest upper left corner) to the highest frequencies (lower right corner) [13].

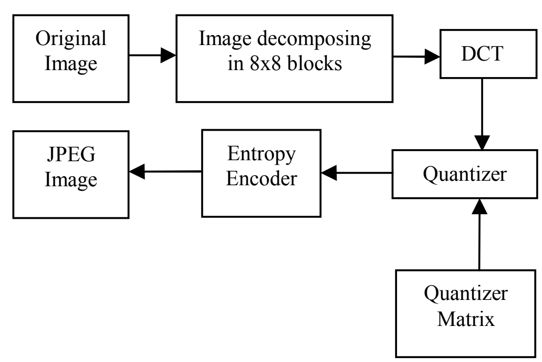


Fig. 1: JPEG Compression Algorithm

III. STREAM CIPHER

RC4 is a stream cipher, symmetric key algorithm. The same algorithm is used for both encryption and decryption as the data stream is simply XORed with the

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

generated key sequence. The key stream is completely independent of the plaintext used. It uses a variable length key from 1 to 256 bit to initialize a 256-bit state table. The state table is used for subsequent generation of pseudo-random bits and then to generate a pseudo-random stream which is XORed with the plaintext to give the ciphertext.

The algorithm can be broken into two stages: initialization, and operation. In the initialization stage the 256-bit state table, **S** is populated, using the key, **K** as a seed. Once the state table is setup, it continues to be modified in a regular pattern as data is encrypted.

The initialization process can be summarized by the pseudo-code:

```
j = 0;
for i = 0 to 255:
S[i] = i;
for i = 0 to 255:
j = (j + S[i] + K[i]) mod 256;
swap S[i] and S[j];
```

It is important to notice here the swapping of the locations of the numbers 0 to 255 (each of which occurs only once) in the state table. The values of the state table are provided. Once the initialization process is completed, the operation process may be summarized as shown by the pseudo code below;

```
i = j = 0;
for (k = 0 to N−1) {
i = (i + 1) mod 256;
j = (j + S[i]) mod 256;
swap S[i] and S[j];
pr = S[ (S[i] + S[j]) mod 256]
output M[k] XOR pr
}
```

Where M [0..N−1] is the input message consisting of N bits. This algorithm produces a stream of pseudo-random values. The input stream is XORed with these values, bit by bit. The encryption and decryption process is the same as the data stream is simply XORed with the generated key sequence.

IV. THE PROPOSED METHOD

In the proposed method, a comparative study of selective image encryption using DCT with Stream Cipher is done. In the DCT method, the basic idea is to decompose the image into 8×8 blocks and these blocks are transformed from the spatial domain to the frequency domain by the DCT. Then, the DCT coefficients correlated to the lower frequencies of the image block are encrypted using the RC4 Stream Cipher. The concept behind encrypting only some selective DCT coefficients (the coefficients [0,0], [0,1], [0,2], [1,0], [2,0], [1,1]) is based on the fact that the image details are situated in the lower frequencies and the human is most sensitive to the lower frequencies than to the higher frequencies. An extra security has been provided to the resulted encrypted blocks by shuffling the resulted blocks using the Shuffling Algorithm. Fig. 4 shows the general block diagram of the proposed method of selective image encryption.

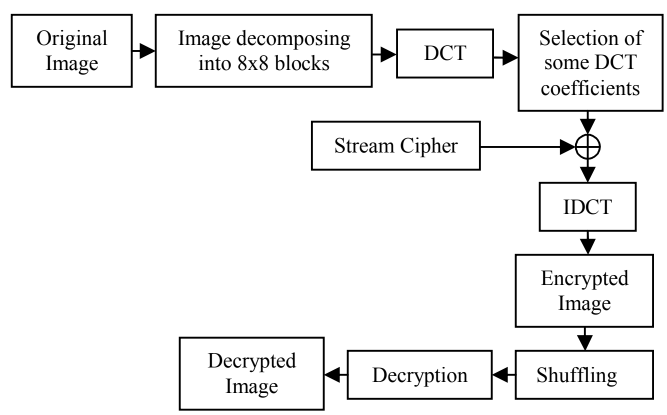


Fig. 4: Block Diagram of the Proposed Method

In the following, the encryption, decryption and shuffling of the images are illustrated.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Algorithm to Encrypt Image

Input : Target Image to be encrypted and the stream RC4 Key values.

Output: Encrypted Image

Begin

Step 1: Read the image header, save the height of the image in variable height & the width in variable width and save the body image in an array imagbody.

Step 2: Obtain how many blocks exist in an image row and how many ones in the column, by dividing the width and height of the image by N, where N is equal to 8 (the required block size).

NoRowB = Image Height / N;

NoColB = Image Width / N;

Step 3: For all blocks in the image perform the following:

- Get_block (row_no, col_no)
- Perform a DCT on the block and save the resulted coefficients in an array.
- Round the selected coefficients, convert the selected coefficients to 11 bits; the 12^{th} bit is used to save the sign of the coefficient.
- Encrypt the selected coefficients by XORing the generated bit stream from the RC4 + Key with the coefficient bits, the sign bit of the selected coefficients will not be encrypted.
- Perform an Inverse Discrete Cosine Transform (IDCT) and get the new block values and the resulted values could be positive or negative values due to the encryption step.

Step 4: Apply the proposed shuffling algorithm on the resulted blocks to obtain the encrypted image.

End

Algorithm to Decrypt Image

Input : Target Image to be decrypted and the Encryption Key

Output: Original Image

Begin

Step 1: Read the image header, save the height of the image in variable height & the width in variable width and save the body image in an array imagbody.

Step 2: Obtain how many blocks exist in an image row and how many ones in the column, by dividing the width and height of the image by N, where N is equal to 8 (the required block size).

NoRowB = Image Height / N;

NoColB = Image Width / N;

Step 3: For all blocks in the image perform the following:

- Get_block (row_no, col_no)
- Perform a DCT on the block and save the resulted values in an array.
- Round the selected coefficients, convert the selected coefficients to 11 bits; the 12th bit is used to save the sign of the coefficient.
- Decrypt the resulted bits by using the generated bit stream from the RC4 + Key, by performing an XOR operation, the sign bit of the selected coefficients will remain.
- Convert the resulted bits into integer values, and join the sign (from the step above) with each integer, if the coefficient is negative multiply it by -1.
- Perform an IDCT and get the new blocks.

Step 4: Reshuffle the block, since the shuffling algorithm generates the same row and column numbers to return the shuffled blocks into their original locations.

Step 5: Reconstruct the image to get the original Image.

End

Shuffling Algorithm

Input : Key, number of blocks in the row (**NoRows**), number of blocks in the column (**NoCols**) and the resulted encrypted image saved in an array.

Output: A new shuffled image

Begin

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

```
        for i = 0 to (NoRows × NoCols)
        NewVal [i] = (Key × i) mod (NoRows × NoCols)
        endfor
        k = 0
        for i = 0 to (NoRows × NoCols)
        MoveBlock (ImageBlk (NewVal [i]), ImageBlk [k])
        k++
        endfor
End
```

V. EXPERIMENTAL RESULTS

The performance analysis of selective image encryption using DCT with Stream Cipher is measured using the Peak Signal to Noise Ratio (PSNR), histogram analysis and entropy. Fig. 5.1 shows the Original Image used in the DCT method. Fig. 5.2 shows the Selective Encryption of the original image. The Encrypted Image after applying the shuffling algorithm is shown in Fig. 5.3 and in Fig 5.4, the Decrypted Image is shown. Here, the number of coefficients/ block selected is 6.

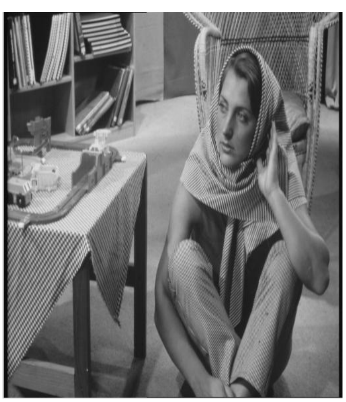


Fig. 5.1: Original Image

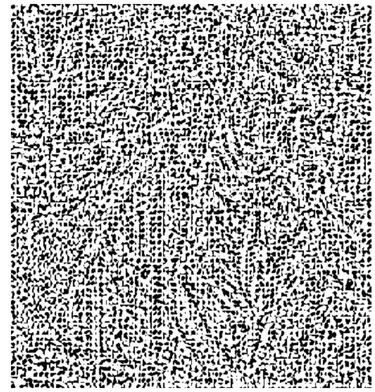


Fig. 5.2: Selective Encryption

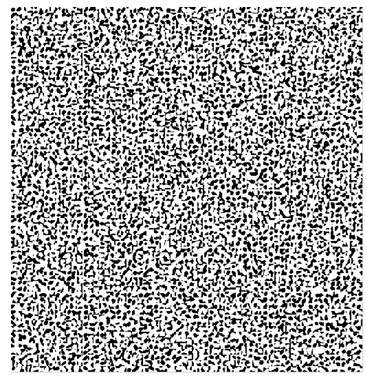


Fig. 5.3: Encrypted Image
(After Shuffling)

Fig. 5.4: Decrypted Image

Table I shows the Performance Analysis of the DCT method. When the number of coefficients/block selected is 6, we obtain a lower PSNR in the case of Encrypted Image and a higher PSNR in the case of Decrypted Image. Higher PSNR value shows a better quality of the image.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Table I

Performance Analysis of DCT Method

Number of coefficients/ block	PSNR of Encrypted Image	PSNR of Decrypted Image
3	32.5619	52.9979
6	29.2989	75.0756
10	29.1731	74.9530
15	28.9983	74.8003

Table II shows the performance analysis of encrypted and decrypted images in terms of PSNR when tested with different test images of size 512×512. A lower PSNR in obtained in the case of Encrypted Images and a higher PSNR is obtained in the case of Decrypted Image. Higher PSNR value shows better quality of the images.

Table II

Performance Analysis of DCT Method with different test images

Test Images	PSNR of Encrypted Image	PSNR of Decrypted Image
Barbara	20.5784	65.6641
House	20.7056	65.4996
Lena	20.8768	65.5393
Airplane	20.6219	65.4215
Baboon	20.7354	65.3072

To demonstrate that our proposed algorithm has strong resistance to statistical attacks, test is carried out on the histogram of enciphered image. Several gray-scale images of size 512×512 are selected for this purpose and their histograms are compared with their corresponding ciphered image. One typical example is shown below. The histogram of the original image contains large spikes as shown in Fig. 5.5 but the histogram of the cipher image as shown in Fig. 5.6, is more uniform. It is clear that the histogram of the encrypted image is, significantly different from the respective histogram of the original image and bears no statistical resemblance to the plain image. Hence statistical attack on the proposed image encryption procedure is difficult.

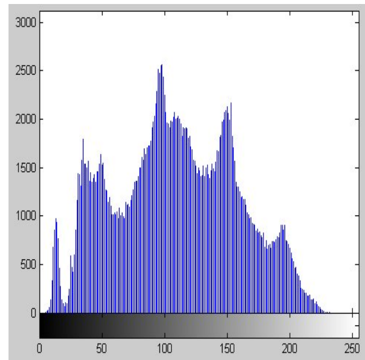


Fig. 5.5: Histogram of Original Image

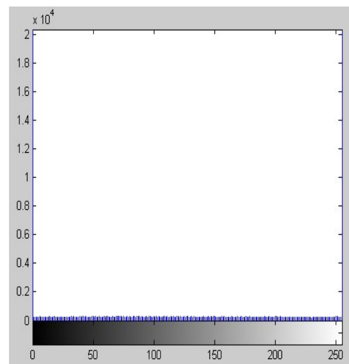


Fig. 5.6 Histogram of Encrypted Image (after shuffling)

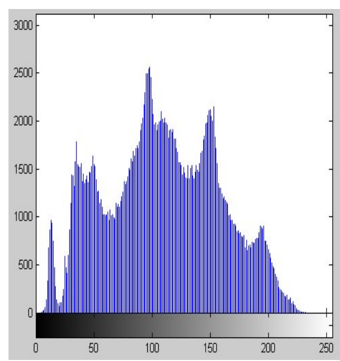


Fig. 5.7 Histogram of Decrypted Image

Entropy is a statistical measure of randomness. Table III shows the entropy of different test images of size 512×512.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Table III

Entropy of different test images

Test Images	Entropy of Encrypted Image
Barbara	3.9693
House	3.9525
Lena	3.9654
Airplane	3.9571
Baboon	3.9470

VI. CONCLUSION

The proposed encryption method uses the Selective Encryption approach where the DC coefficients and some selective AC coefficients are encrypted, hence the DC coefficients carry important visual information, and it's difficult to predict the selective AC coefficients, this give a high level of security in comparison with methods mentioned above. The algorithm will not encrypt bit by bit the whole image but only selective DCT coefficients will be encrypted, and extra security has been added to the resulted encrypted blocks by using Shuffling method. The algorithm considered as a fast image encryption algorithm, due to the selective encryption of certain portion of the image (the DC and some AC coefficients). PSNR values of the encrypted images are low and are resistant to statistical attacks. Hence, better security has been provided.

REFERENCES

[1] Lala Krikor, Sami Baba, Thawar Arif, Zyad Shaaban, "Image Encryption Using DCT and Stream Cipher", European Journal of Scientific Research, ISSN 1450-216X, Vol.32, No.1 (2009), pp.47-57.

[2] Xiliang Liu, "Selective Encryption of Multimedia Content in Distribution Networks: Challenges and New Directions", Proceedings of Communications, Internet, and Information Technology (CIIT 2003), Scottsdale, AZ, USA, Nov. 2003.

[3] M. Van Droogenbroeck and R. Benedett, "Techniques for a Selective Encryption of Uncompressed and Compressed Images", in Proceedings of Advanced Concepts for Intelligent Vision Systems (ACIVS) 2002, Ghent, Belgium, Sept. 2002.

[4] M. M. Fisch, H. Stgner, and A. Uhl, "Layered Encryption Techniques for DCT-Coded Visual Data", in European Signal Processing Conference (EUSIPCO) 2004, Vienna, Austria, Sep., 2004.

[5] Rodrigues, J.M. Puech, W. Bors, A.G. "Selective Encryption of Human Skin in JPEG Images", IEEE International Conference on Image Processing, 2006.

[6] Puech, W.; Rodrigues, J.M.; Bors, A.G., "Analysis and Cryptanalysis of a Selective Encryption Method for JPEG Images", Eighth International Workshop on Image Analysis for Multimedia Interactive Services, 2007. WIAMIS07, June 2007.

[7] L. Tang., "Methods for Encrypting and Decrypting MPEG Video Data Efficiently", In Proc. ACM Multimedia, volume 3, pages 219–229, 1996.

[8] H. Cheng and X. Li., "Partial Encryption of Compressed Images and Videos", IEEE Trans. on Signal Processing, 48(8):2439–2445, Aug. 2000.

[9] R. Lukac, K. Plataniotis, "Bit-Level Based Secret Sharing for Image Encryption", Pattern Recognition, 38(5):767–772, May 2005.

[10] C. Wu, C. Kuo. "Design of Integrated Multimedia Compression and Encryption Systems", IEEE Trans. on Multimedia, 7(5):828–839, Oct. 2005.

[11] M. V. Droogenbroeck, R. Benedett, "Techniques for a Selective Encryption of Uncompressed and Compressed Images", In Proc. of Advanced Concepts for Intelligent Vision Systems (ACIVS) 2002, Ghent, Belgium, pages 90–97, Sept. 2002.

[12] C. Coconu, V. Stoica, F. Ionescu, and D. Profeta, "Distributed Implementation of Discrete Cosine Transform Algorithm on a Network of Workstations", Proceedings of the International Workshop Trends & Recent Achievements in IT, Romania, pp. 116-121, May 2002.

[13] JPEG, jpeg.org.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

(IJCSIS) International Journal of Computer Science and Information Security,
Vol. 8, No. 4, July 2010

Adaptive Background Estimation and object detection applying in Automated visual surveillance

M.Sankari
Department of Computer Applications,
Nehru Institute of Engineering and Technology,
Coimbatore, INDIA.
sankarim2@gmail.com

C. Meena
Head, Computer centre,
Avinashilingam University,
Coimbatore, INDIA.
cccmeena@gmail.com

***Abstract*— Automated visual surveillance is currently a hot topic in computer vision research. A common approach is to perform background subtraction, which identifies moving objects from the sequence of video frames that differs significantly from a background model. As a basic, the background image must be a representation of the scene with no moving objects and must be kept regularly updated. There are many challenges in developing a good background subtraction algorithm. We have proposed a methodology to perform background subtraction from moving vehicles in traffic video sequences that combines statistical assumptions of moving objects using the previous frames. It is necessary to update the background image frequently in order to guarantee reliability of the motion detection. For that, a binary moving objects hypothesis mask is constructed to classify any group of lattices as being from a moving object based on the optimal threshold. Then, the new incoming information is integrated into the current background image using a Kalman filter. In order to improve the performance, it is necessary to perform post- processing. It has been accomplished by shadow and noise removal algorithms operating at the lattice which identifies object-level elements. The results of post-processing can be used to detect object more efficiently. Experimental results and comparisons using real data demonstrate the superiority of the proposed approach which has achieved an average accuracy of 92% on completely novel test images.**

Keywords- Background subtraction; Background updation; Binary segmentation mask; Kalman filter; Noise remova; Shadow removal; Traffic video sequences.

I. Introduction

As computer vision begins to address the visual interpretation of many problems in applications such as surveillance and monitoring are becoming more relevant. Our main goal is to identify the object from the multi model background. For that we need to detect and extract the foreground object from the background image. Once objects are detected, the further processing for tracking and activity is limited in the corresponding regions of the image. Several factors make on-road vehicle detection very challenging including variability in scale, location, orientation, and pose. Vehicles, for example, come into view with different speeds and may vary in shape, size, and color. Vehicle appearance depends on its pose and is affected by nearby objects. In-class variability, occlusion, and lighting conditions also change the overall appearance of vehicles. Region along the road changes continuously while the lighting conditions depend on the time of the day and the weather.

We present a system for detecting and tracking vehicles in surveillance video which uses a simple motion model to determine salient regions in a sequence of video frames. Similar regions are associated between frames and grouped to form the background. The entire process is automatic and uses computation time that scales according to the size of the input Video sequence. We consider image/video segmentation with initial background subtraction, object tracking, and vehicle counting, in the domain of traffic monitoring over an intersection.

The remainder of the paper is organized as follows: Section II gives the overview of the related work. Section III describes the architecture and modeling of proposed methodology for background elimination and object detection. Implementation and performance are analyzed in section IV. Section V contains the concluding remarks and future work.

II. Overview of the Related Work

Many works have been proposed in the literature as a solution to an efficient and reliable background subtraction. To detect moving objects in a dynamic scene, adaptive background subtraction techniques have been developed [1] [2] [3]. Adaptive Gaussian mixtures are commonly chosen for their analytical representation and theoretical foundations. For these reasons, they have been employed in real-time surveillance systems for background subtraction [4] [5] and object tracking [6]. For foreground analysis [7] [8], a method for foreground analysis was proposed for moving object, shadow, and ghost by combining the motion information. The computation cost is relatively expensive for real-time video surveillance systems because of the computation of optical flow. In [9], a work has presented on a novel background subtraction algorithm that is capable of detecting objects of interest while all pixels are in motion. Background subtraction technique is mostly used for motion pictures to segment the foreground object by most of the researchers [10] [11]. Liyuan Li, et al. [12] proposed foreground object detection through foreground and background classifications under bayesian

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

framework. In addition, moving object segmentation with background suppression is affected by the problem of shadows [6] [13]. Indeed, the moving object detection do not classify shadows as belonging to foreground objects since the appearance and geometrical properties of the object can be distorted which, in turn, affects many subsequent tasks such as object classification and the assessment of moving object position. In this paper, we propose a novel simple method that exploits all these features, combining them so as to efficiently provide detection of moving objects, ghosts, and shadows. The main contribution of this proposal is the integration of knowledge of detected objects, shadows, and ghosts in the segmentation process to enhance both object segmentation and background update. The resulting method proves to be accurate and reactive and, at the same time, fast and flexible in the applications.

III. PROPOSED WORK

Block diagram of the algorithm used in proposed system is described in Fig.1. The proposed Automated Video Surveillance system seeks to automatically identify people, objects, or events of interest in variety of environment. Typically, these systems consist of stationary cameras placed in highways. These cameras are integrated with, intelligent computer systems that perform preprocessing operation from the captured video images and notify human operators or trigger control process. The objective of this real-time motion detection and tracking algorithm is to provide low-level functionality for building higher-level recognition capabilities.

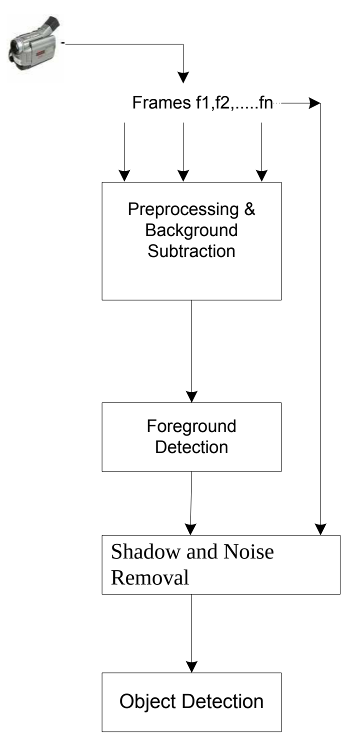


Figure 1. Block diagram of the algorithm.

A. Preprocessing

Preprocessing is the key step and the starting point for image analysis, due to the wide diversity of resolution, image format, sampling models, and illumination techniques that are used during acquisition. In our method, preprocessing step was done by statistical method using adaptive median filter. The resultant frames are then utilized as an input for the background subtraction module. In Fig.2, image *I(x,y)* at time *t* is shown in Fig.3 , the background image *B(x,y)* at time *t* is shown.

Figure 2. Image at time *t*: *I (x; y; t).*

Figure 3. Image at time *t*: *B (x; y; t).*

To get the new estimated background, the following steps were used for calculation.

Step 1: Estimate the background at time t using adaptive median filter method.

Step 2: Subtract the estimated background from the input frame.

Step 3: Apply a threshold, to the absolute difference to get the binary moving objects hypothesis mask.

Assuming that the background is more likely to appear in a scene, we can use the median of the previous n frames as the background model

$$B(x,y,t) = Median((I(x,y,t-i)), \qquad (1)$$

$$| I(x,y,t) = Median(I(x,y,t-i) |> \varphi, \qquad (2)$$

where $i \in \{0,1,2,\ldots,n-1\}$

B. Foreground Detection

In this module estimated background and foreground mask images are used as an input for further processing .Thus, we use grayscale image sequences as input. Elements of the scene and the sizes of the traffic objects (vehicles and pedestrians) are unknown. The Foreground detection is done by using accumulative difference method, which is change-detection based on subtraction of a background image. It is necessary to update the background image frequently in order to guarantee reliable object detection. The basic idea in background

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

adaptation is to integrate the new incoming information into the current background image using a Kalman filter:

$$B_{(t+1)} = B_t + [a_1 * (1 - M_t + a_2 * M_t] D_t \,, \tag{3}$$

where B_t represents the background model at time t, D_t is the difference between the present frame and the background model, and M_t is the binary moving objects hypothesis mask. The gain a_1 *and* a_2 are based on an estimate of the rate of change of the background. The larger it is, the faster new changes in the scene are updated to the background frame. In our approach, $a_1 = 0.1$ *and* $a_2 = 0.01$, they are kept small and the update process based on Eq.(3) is only intended for adapting to slow changes in overall lighting.

$$M_t(x) = \begin{cases} 1, & if \; |D_t(x)| > T_t \\ 0, & Otherwise \end{cases} \,. \tag{4}$$

Foreground detection is started by computing a pixel based absolute difference between each incoming frame and an adaptive background frame B_t . The pixels are assumed to contain motion if the absolute differences exceed a predefined threshold level. As a result, a binary image is formed where active pixels are labeled with a "1" and non-active ones with a "0". With the updated background image strategy using Kalman filter, we get the better foreground detection result. This is a simple, but efficient method to monitor the changes in active during a few consecutive frames. Those pixels which tend to change their activity frequently are masked out from the binary image representing the foreground detection result**.**

C. Shadow Removal

Shadows appear as surface features, when in fact they are caused by the interaction between light and objects. This may lead to problems in scene understanding, object segmentation, tracking, recognition, etc. Because of the undesirable effects of shadows on image analysis, much attention was paid to the area of shadow detection and removal over the past decades and covered many specific applications such as traffic surveillance. In this paper, 8-neighborhood gray clustering method is used to define the precise shadow and remove it. The mean clustering threshold and the initial cluster seed of the gray are calculated by the following equations.

$$T_i = (1/3)\max(G(x,y) - u_i)^2 \,, \tag{5}$$

where $G(x, y)$ is a gray value of the pixel in $I(x, y)$, u_i is the mean of $G(x, y)$. The initial seed C_i locates in the centre of $G(x, y)$.The clustering starts from the seed C_i , and the point P_i is examined in turn. If at least one point in the 8-neighborhood of P_i has been marked as a shadow region, standard deviation of P_i is calculated by the following equation.

$$\delta P_i = (P_i(x, y) - u_i)^2 \,, \tag{6}$$

where $P_i(I(x, y))$ is the gray value of I(x,y). If $\delta P_i < T$, P_i must be shadow point; otherwise, the point need not be marked. The point P_i is checked constantly until no new point is marked. At last all the marked shadow points are removed.

D. Noise removal

Usually due to the camera noise and irregular object motion, there always exist some noise regions both in the object and background region. Moreover, the object boundaries are also not very smooth. Hence a post processing technique is applied on the foreground image. In order to remove the noise wiener low-pass filters a grayscale image that has been degraded by constant power additive noise. It is based on an adaptive statistics estimated from a local neighborhood of each pixel.

E. Detection

After post-processing, the image is compared with the one of the original frames (usually, the first frame). If the pixels are less than certain threshold, then they are ignored. Otherwise, they are replaced by the pixels of original image. This resulting image will be consisting of the moving object ignoring the background and hence satisfying our requirement.

IV. IMPLEMENTATION AND PERFORMANCE ANALYSIS

This system was implemented on an Intel Pentium IV 280 GHz PC. We have tested the system on image sequences on different scenarios like traffic junction intersection, highways etc. Real life traffic video sequence are used to demonstrate the vehicle tracking from traffic video sequences using the proposed framework. All the videos chosen for vehicle tracking have same light intensity and have been taken during day time. We convert the colour video frames to gray scale images.

Automatic monitoring visual surveillance system implementation needs to detect vehicles using automatic background extraction. Background subtraction is the main step for vehicle detection. Fig. 4 shows number of successive frames that are used to extract the background. Digital camera used to take shots. The camera placed over the highway directly. It shots eight frames per second.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Figure 4. Number of successive frames that are used for preprocessing.

Estimated background using Median Filter for n = 12. This is shown in Fig.5

Figure 5. Estimated background.

After apply a threshold, to the absolute difference we got the binary moving objects hypothesis mask which is shown in Fig. 6

Figure 6. Binary moving objects hypothesis mask.

Fig.7 shows the foreground detected objects obtained after background subtracted, shadow and noise. The automatic background extraction results are very good and promising. The most effective parameters that are playing a main role for automatic background extraction are the threshold level. This threshold is used to extract the moving vehicles from the background. Matlab built-in function has been employed for the evaluation of the threshold.

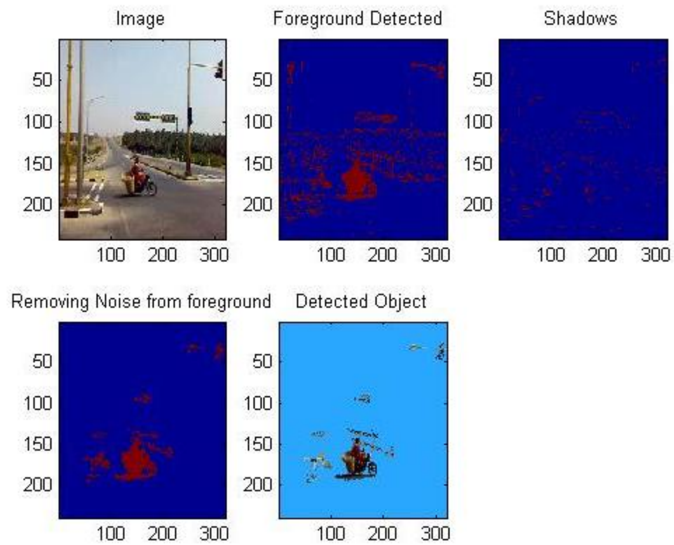


Figure 7. Foreground object detection process.

Table I gives vehicle detection results. The background is subtracted from the current image then the resulted image is filtered to get moving vehicles only. By using this technique most of vehicles are detected. Moving vehicles are detected easily after background is subtracted. Performance analysis is shown in Fig.8.

TABLE I. The results for vehicle detection.

Type of Vehicle	Actual Number of vehicles	Detected Vehicles	Rate %
Car	20	19	95
Motor cycle	10	9	90
Bus	15	14	93
Lorry	10	8	80
Truck	13	12	92

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

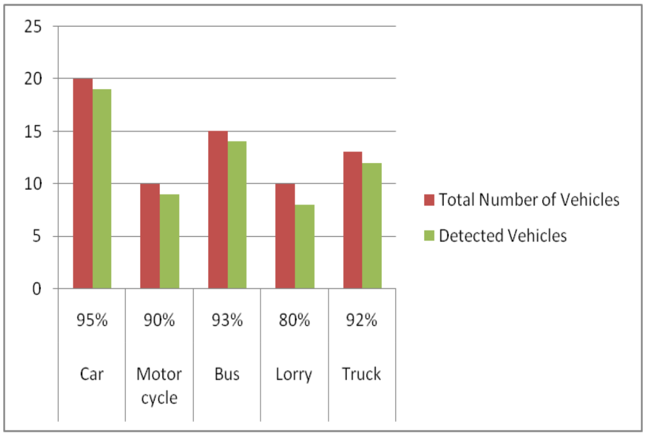


Figure.8 Performance analysis.

V. CONCLUSION

The experimental results of applying this approach lead to detect moving vehicles efficiently. This approach gives promising and effective results where vehicle detection rate was around 94%. In this approach the advantages of background subtraction and edge detection are used. It could be improved and used as a basis for automatic traffic monitoring. Miss detection resulted from occluding large vehicles the small ones and the far moving vehicles that appear as a point in the image. These difficulties could be solved in the future work.

ACKNOWLEDGMENT

Authors thank their family members and children for their continuous support and consent encouragement to do this research work successfully.

REFERENCES

[1] M. Cristani, M. Bicegi, and V. Murino, “Integrated Region- and Pixel-based Approach to Background Modeling”, Proceedings of the MOTION, 2002.
[2] H. Eng, J. Wang, A. Kam, and W. Yau, “Novel Regionbased Modeling for Human Detection within High Dynamic Aquatic Environment,” Proceedings on CVPR, 2004.
[3] P. KaewTraKulPong and R. Bowden, “An Improved Adaptive Background Mixture Model for Real-time Tracking with Shadow Detection,” In Proc. 2nd European Workshop on Advanced Video Based Surveillance Systems, 2001.
[4] M. Harville, G. Gordon, and J. Woodfill, “Foreground Segmentation Using Adaptive Mixture Models in Color and Depth,” Proc. ICCV Workshop Detection and Recognition of Events in Video, July 2001.
[5] C. Stauffer and W.E.L. Grimson, “Adaptive Background Mixture Models for Real-Time Tracking,” Proc. Conf. Computer Vision and Pattern Recognition, vol. 2, pp. 246-252, June 1999.
[6] S.J. McKenna, Y. Raja, and S. Gong, “Object Tracking Using Adaptive Color Mixture Models,” Proc. Asian Conf. Computer Vision, vol. 1, pp. 615-622, Jan. 1998.
[7] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, “Detecting Moving Objects, Ghosts, and Shadows in Video Streams,” IEEE Trans. on PAMI, 25: (10), October 2003.
[8] Connell, J., “Detection and Tracking in the IBM PeopleVision System,” IEEE ICME, June 2004.
[9] Dongxiang Zhou, Hong Zhang and Nilanjan Ray, “Texture Based Background Subtraction,” *Proc. IEEE Int. Conf. on Information and Automation*, Zhangjiajie, China,
pp. 601-605, 2008.
[10] A. McIvor, “Background subtraction techniques,” in *Proc. Image Video Computing*, pp. 147-153, 2000.
[11] Y. Ivanov, A. Bobick, and J. Liu, “Fast lighting independent background subtraction,” *Int. Journal on Comp. Vision*, vol. 37, no. 2, pp. 199–207, Jun. 2000.
[12] Liyuan Li, Weimin Huang, Irene Yu-Hua Gu and Qi Tian, “Statistical Modeling of Complex Backgrounds for Foreground Object Detection,” *IEEE Transactions on Image Proc.*, vol. 13, no. 11, pp. 1459-1472, Nov 2004.
[13] A. Elgammal, D. Harwood, and L.S. Davis, “Non-Parametric Model for Background Subtraction,” Proc. IEEE Int’l Conf. Computer Vision ’99 FRAME-RATE Workshop, 1999.

AUTHORS PROFILE

Mrs. M. Sankari received her B.Sc. and M.Sc. degrees in Computer science from Bharathidasan University in 1988 and 1990, respectively. She has completed her Master of Philosophy degree in Computer science from Regional Engineering College, Trichy in 2000. Presently, she is a Head of the department of MCA at NIET and pursuing her doctorate degree in computer science at Avinashilingam University, Coimbatore, India. She has published various technical papers at IEEE conferences. Her field of research includes Computer vision, Pattern recognition, Analysis of algorithms, Data structure, Computer graphics and multimedia.

Dr. C. Meena received her B.Sc (Physics), and Master of Computer Applications degrees from Madurai Kamaraj University in 1987 and 1990, respectively. She has completed her Ph.D. degree in Computer science from Bharathiyar University in 2006. Presently, she is Head, Computer Centre at Avinashilingam University, Coimbatore, India. She is guiding for funded Research Project in UGC and Naval Research Board. She has presented international/national journals. She has published various technical papers at IEEE conferences and national level conferences. Her field of research includes Image processing, Computer vision and Pattern recognition.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Securing Web Communication with Quantum Cryptography

R.K.Pateriya[1], R.K. Baghel[2], Anupriya Gupta[3]
[1]Associate Professor, Department of Information Technology
[2]Associate Professor, Department of Electronics Engineering
[3]M.Tech (Information Security) Scholar, Department of Computer Science & Engineering
Maulana Azad National Institute of Technology, Bhopal, India
Emails: pateriyark@gmail.com , baghel.rk@gmail.com, er.naina@gmail.com

Abstract—***The problem of transmitting secret messages securely between two parties is very old one. Human imagination has come up with clever ways of overcoming the difficulties associated with this problem, in particular preventing a malevolent eavesdropper from obtaining information about the secret message exchanged over the communication channel. Now a days internet security is most important issue because a large number of people depends on online transaction. During recent year quantum cryptography has been the object of a strong activity and is now extending its activity into various areas. Quantum cryptography is now a days widely used for communicating secret data between two authenticated parties. It has great potential to become the key technology for securing confidentiality and privacy of communication and thus to become the driver for the success of a series of web services in the field of e-governance, e-commerce, e-health, transmission of biometric data etc. The main problem with quantum cryptography is to find the initial raw key. This problem is discussed in this paper and a method is proposed which uses quantum cryptography in SSL/TLS server for securing web communication.***

Keywords- BB84 protocol , Random key generation, QKD.

I. INTRODUCTION

Quantum cryptography is an approach to securing communications by applying the phenomena of quantum physics. Unlike traditional classical cryptography, which uses mathematical techniques to restrict eavesdroppers, quantum cryptography is focused on the physics of information. Unlike the public key cryptosystem the security of QKD is provable and can not be compromised. Although QKD is not very practical but it has been the object of intensive research activities and of rapid progress. It is only useful for short distance but with sufficient technical improvement it will become possible to implement QKD over large distances. Unconditional security of QKD lies in the principal of Heisenberg uncertainty and principal of photon polarization. According to the Current status of security of classical cryptosystem in relation to quantum cryptography[1].

Cryptosystem	*Broken by Quantum algorithm*
RSA public key encryption	Broken
Diffie-Hellman key-exchange	Broken
Elliptic curve cryptography	Broken

TABLE I. SECURITY STATUS OF CRYPTOSYSTEM

The main protocol utilized in QKD is BB84 protocol. It consist of four stages [2] .

A. *Stages Utilized in BB84 Protocol*

Raw key generation: Transmission of randomly encoded single photon stream over the quantum channel from sender to receiver to establish initial raw key. The sender will maintain the temporary database of each photon sent.

Sifting: Here receiver will send a list of photons detected and their basis but not their value back to the sender over the same channel. photons are measured with three basis horizontal, vertical, diagonal and only one basis can be applied to one photon as it is measured once.

Reconciliation: This phase mainly deals with the correction of errors. This process requires a number of communication between sender and receiver over classical channel and because of this, the size of key is reduced than sifted key.

Privacy amplification: In this step no communication is needed between sender and receiver. Using reconciled set of bits a new smaller set of bits is computed.

B. QKD based on three levels of complexity

QKD is a cryptographic primitive which is used for various purposes of increasing complexity. These purposes can be classified according to the three levels of complexity. These levels are equivalent to the first three layers of the OSI model.

- The first level is key establishment between two user shared by each . Quantum Key Agreement, falls in the category of physical layer security cryptographic primitives.
- The second level is two user secure payload transmission built on top of key establishment scheme. These are link layer security cryptographic primitives.
- The third level is key distribution over a global network composed of multiple users. These are network layer security cryptographic primitives.

II. PREVIOUS WORK

Quantum computing is a rapidly growing field of research that applies concepts of quantum physics to building more efficient computers. Although only rudimentary quantum computers have been built so far, many researchers believe that quantum computing has great potential. In recent years, there has been extensive studies about the possibilities offered by quantum computation to cryptology. From the point of view of quantum computing researchers, after the advent of high power quantum computers, conventional cryptography may be no longer secure. Cryptanalysis tasks would be dramatically accelerated with the help of quantum computers, if such computers are ever build. Till now quantum cryptography is implemented successfully in optical fiber communication but large distances is still a problem to be solved. There are drawbacks in the existing techniques like initially deciding a fixed length of key is a problem. As distance increases key size varies and error rate also increases. Presently sequential key generation is being used as a result of which if any intermediate sample of data get corrupted than all preceding sample get affected and because of this error rate increases. Following method is used for key generation in existing techniques [3]

$$\tau \leq \begin{cases} \log_2\left(1+4e-4e^2\right) & \text{for } e \leq 1/2 \\ 1 & \text{for } 1/2 \leq e \end{cases}$$

$$k = n\{1-\tau+f\,[e]\,(e\log_2(e)+(1-e)\log_2(1-e))\}$$

Where n=length of raw key, e=error rate and k=final key that is obtained after error correction.

III. PROPOSED MODIFIED SSL/TLS RECORD PROTOCOL WITH QUANTUM CRYPTOGRAPHY

Here a modification is proposed in the Transport Layer Record Protocol of SSL/TLS using quantum cryptography key distribution. In this modification three parts of QKD are occurred first one is QKD setting protocol for the bias setting and channel establishment . Second is one hand shake protocol for communication for the client and server and finally amplification of encrypted message for authentication purpose.

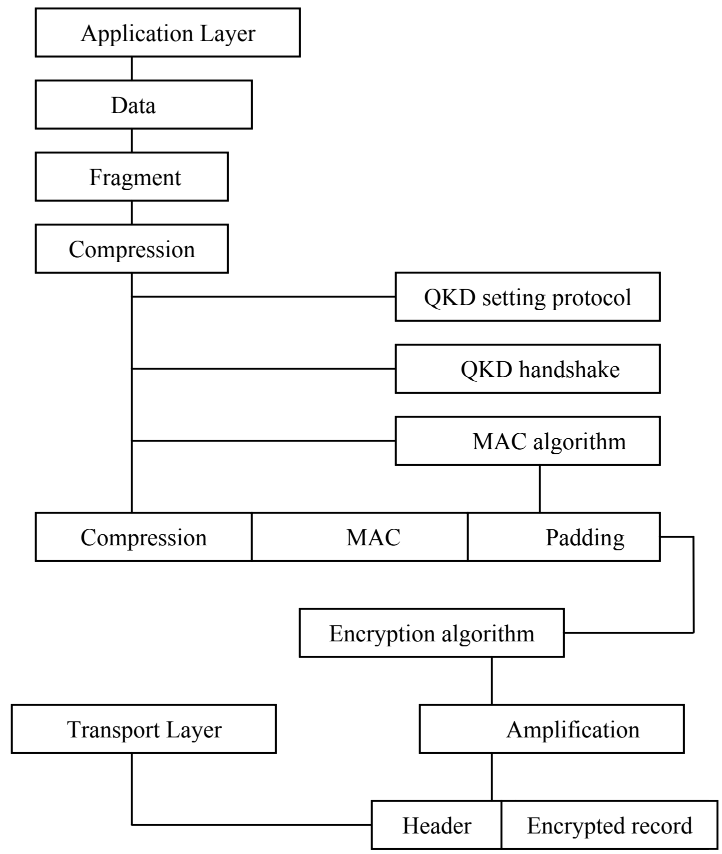


Figure 1. SSL/TLS Record Protocol

Our proposal is the random key generation for getting the initial key of fixed length. With this error rate will be reduced because corruption of intermediate sample of data will not affect the whole process of key generation. The Basic problem of QKD is to find the initial key length will also be removed remove by our proposed technique . As soon as raw key get fixed, base announcement is done and channel get established for further communication. We will simulate QKD in SSL server and make an error generator for estimating the error to know which bit is get corrupted and this data will be used in our random key generation scheme using monte carlo method [4].

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

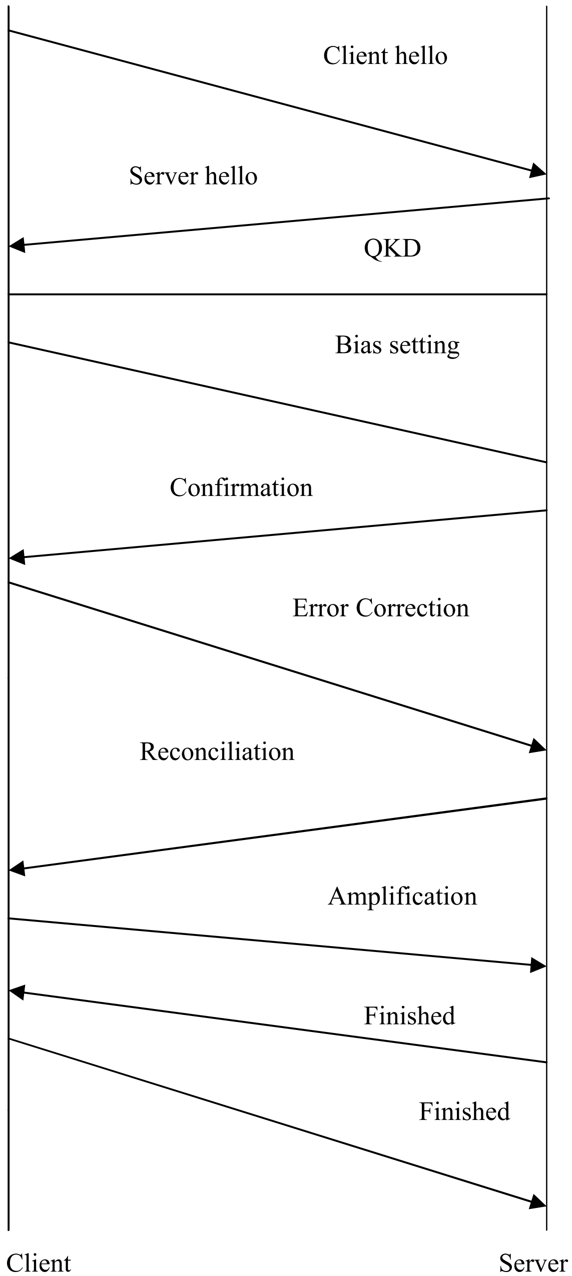


Figure 2. SSL Using QKD

Monte Carlo Key Generation Method

In this we are integrating new concept of random sifting of key length using random number generator.

Consider a random variable X having probability density function $f_x(x)$ which is greater than zero, than expected value of function g of x is

$$E(g(x))= \sum_{x \varepsilon X} g(x)\, f_x(x) \qquad (1)$$

If x is discrete and

$$Eg(x)=\int_{x \varepsilon X} g(x)\, f_x(x)dx \qquad (2)$$

Now the value of threshold of key shift is

$$\tau \leq \begin{cases} \log_2 1+4e-4e^2 & \text{for } e \leq ½ \\ 1 & \text{for } 1/2 \leq e \end{cases}$$

After modification

$$E(\tau)= \int_{x \varepsilon e} g\,(\log_2 1+4e-4e^2)\, tdt$$

Now the final key length is

$K=\{1- E(\tau)+f(e)$

Where $f(e)=e\log_2(e)+(1-e)\log_2(1-e)$

IV. CONCLUSION AND FUTURE WORK

In this paper a new technique is proposed in which random key is generated for getting the initial key of fixed length Which will reduce the error rate because corruption of intermediate sample of data will not affect the whole process of key generation. The future work will be to evaluate the performance of key found by new method and analyze the results by plotting the graph between error and key for both old and new key.

ACKNOWLEDGMENT

The Success of this research work would have been uncertain without the help and guidance of a dedicated group of people in our institute MANIT Bhopal. We would like to express our true and sincere acknowledgements as the appreciation for their contributions, encouragement and support. The researchers also wish to express gratitude and warmest appreciation to people, who, in any way have contributed and inspired the researchers

REFERENCES

[1]Sean hallgren and Ulrich vollmer “Quantum computing” Springer Berlin Heidelberg. Pages15-34 February 01, 2009

[2] Jörgen Cederlöf and Jan-Åke Larsson “Security Aspects of the Authentication Used in Quantum Cryptography” IEEE Transactions On Information Theory, Vol. 54, No. 4, April 2008.

[3]Mario Pivk, Christian Kollmitzer “SSL/TLS with Quantum Cryptography” Third International Conference on Quantum, Nano and Micro Technologies, 2009.

[4] Eric C. Anderson“ Monte Carlo Methods and Importance Sampling” Lecture Notes for Stat 578C Statistical Genetics1999

[5] Mohamed Elboukhari, Mostafa Azizi, and Abdelmalek Azizi “Integration of Quantum Key Distribution in the TLS Protocol” IJCSNS International Journal of Computer Science and Network Security, VOL.9 No.12, December 2009

[6] Alan Mink, Sheila Frankel and Ray Perlner" Quantum Key Distribution (QKD) and Commodity Security Protocols: Introduction and Integration" International Journal of Network Security & Its Applications (IJNSA), Vol 1, No 2, July 2009.
[7]. Rajni Goel, Moses Garuba, Anteneh Girma "Research Directions in Quantum Cryptography" International Conference on Information Technology (ITNG'07). Pages 779-784,2007
[8].A.Falahati , Hadi Meshgi "Using Quantum Cryptography for securing Wireless LAN networks"International Conference on Signal Processing Systems. Pages 698-701 , 2009
[9] Stamatios V.Kartalopoulos "Identifying vulnerabilities of quantum cryptography in secure optical data transport" IEEE Military communications Conference Milcom , pages 1-9, 2005.

R K Pateriya M.Tech & B.E. in Computer Science & Engg. and working as Associate Professor in Information Technology Department of MANIT Bhopal . Total 17 Years Teaching Experience (PG & UG). Guided more than twenty M.Tech Thesis . Published several

R K Baghel Working as Associate Professor in Electronics Engineering Department of MANIT Bhopal. PhD in Optical communication, M.Tech in Digital Communication & B.E. in Electronics & Communication Engg. Total 20 Years Teaching Experience (PG & UG). Guided more than twenty M.Tech Thesis. His research area is Low power VLSI & Communication

Anupriya Gupta B.E. in Electronics Engineering. Pursuing M.Tech. in Information Security from MANIT Bhopal.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

A Robust -knowledge guided fusion of clustering Ensembles

Anandhi R J
Research Scholar, Dept of CSE,
Dr MGR University,
Chennai, India
rjanandhi@hotmail.com

Dr Natarajan Subramaniyan
Professor, Dept of ISE
PES Institute of Technology
Bangalore, India
snatarajan44@gmail.com

***Abstract*— Discovering interesting, implicit knowledge and general relationships in geographic information databases is very important to understand and to use the spatial data. Spatial Clustering has been recognized as a primary data mining method for knowledge discovery in spatial databases. In this paper, we have analyzed that by using a guided approach in combining the outputs of the various clusterers, we can reduce the intensive computations and also will result in robust clusters .We have discussed our proposed layered cluster merging technique for spatial datasets and used it in our three-phase clustering combination technique in this paper. At the first level, m heterogeneous ensembles are run against the same spatial data set to generate $B_{1...}B_m$ results. The major challenge in fusion of ensembles is the generation of voting matrix or proximity matrix which is in the order of n^2, where n is the number of data points. This is very expensive both in time and space factors, with respect to spatial datasets. Instead, in our method, we compute a symmetric clusterer compatibility matrix of order (m x m) , where m is the number of clusterers and m <<n, using the cumulative similarity between the clusters of the clusterers. This matrix is used for identifying which two clusterers, if considered for fusion initially, will provide more information gain. As we travel down the layered merge, for every layer, we calculate a factor called Degree of Agreement (DOA), based on the agreed clusterers. Using the updated DOA at every layer, the movement of unresolved, unsettled data elements will be handled at much reduced the computational cost. Added to this advantage, we have pruned the datasets after every (m-1)/2 layers, using the gained knowledge in previous layer. This helps in faster convergence compared to the existing cluster aggregation techniques. The correctness and efficiency of the proposed cluster ensemble algorithm is demonstrated on real world datasets available in UCI data repository.**

Keywords- Clustering ensembles, Spatial Data mining, Degree of Agreement, Cluster Compatibility matrix.

I. INTRODUCTION

With a variety of applications, large amounts of spatial and related non-spatial data are collected and stored in Geographic Information Databases. Spatial Data Mining[1], (i.e., discovering interesting, implicit knowledge and general relationships in large spatial databases) is an important task for the understanding the usage of these spatial data. With the rapid growth in size and number of available databases in commercial, industrial, administrative and other applications, it is necessary and interesting to examine how to extract knowledge automatically from huge amount of data. Very large data sets present a challenge for both humans and machine learning algorithms. Machine learning algorithms can be inundated by the flood of data, and become very slow in knowledge extraction. More over, along with the large amount of data available, there is also a compelling need for producing results *accurately* and *fast*.

Efficiency and scalability are, indeed, the key issues when designing data mining systems for very large data sets. Through the extraction of knowledge in databases, large databases will serve as a rich, reliable source for knowledge generation and verification, the discovered knowledge can be applied to information management, query processing, decision-making, process control and many other applications. Therefore, data mining has been considered as one of the most important topics in databases by many database researchers.

Spatial data describes information related to the space occupied by objects. It consists of 2D or 3D points, polygons etc. or points in some d-dimensional feature space. It can be either discrete or continuous. Discrete spatial data might be a single point in multi-dimensional space while continuous spatial data spans a region of space. This data might consist of medical images or map regions and it can be managed through spatial databases [8].

Clustering [17] is to group analogous elements in a data set in accordance with its similarity such that elements in each cluster are similar, while elements from different clusters are dissimilar. It doesn't require the class label information about the data set because it is inherently a data-driven approach. So, the most interesting and well developed method of manipulating and cleaning spatial data in order to prepare it for spatial data mining analysis is by clustering that has been recognized as a primary data mining method for knowledge discovery in spatial database [4-7].

Clustering fusion is the integration of results from various clustering algorithms using a consensus function to yield stable results. Clustering fusion approaches are receiving increasing attention for their capability of improving clustering performance. At present, the usual operational mechanism for

clustering fusion is the combining of clusterer outputs. One tool for such combining or consolidation of results from a portfolio of individual clustering results is a cluster ensemble [13]. It was shown to be useful in a variety of contexts such as "Quality and Robustness" [3], "Knowledge Reuse" [13,14], and "Distributed Computing" [9].

The rest of the paper is organized as follows. The related work is in section 2. The proposed knowledge guided fusion ensemble technique is in section 3. In section 4, we present experimental test platform and results with discussion. Finally, we conclude with a summary and our planned future work in this area of research.

II. RELATED WORK

A. Litereature on Clustering Algorithms

Many clustering algorithms have been developed and they can be roughly classified into hierarchical approaches and non-hierarchical approaches. Non-hierarchical approaches can also be divided into four categories; partitioning methods, density-based methods, grid-based methods, and model-based methods. Hierarchical algorithms can be further divided to agglomerative and divisive algorithms, corresponding to bottom-up and top-down strategies, to build a hierarchical clustering tree.

Spatial data mining or knowledge discovery in spatial databases refers to the extraction, from spatial databases, of implicit knowledge, spatial relations, or other patterns that are not explicitly stored [8, 10]. The large size and high dimensionality of spatial data make the complex patterns that lurk in the data hard to find. It is expected that the coming years will witness very large number of objects that are location enabled to varying degrees. Spatial clustering [8] has been used as an important process in the areas such as geographic analysis, exploring data from sensor networks, traffic control, and environmental studies. Spatial data clustering has been identified as an important technique for many applications and several techniques have been proposed over the past decade based on density-based strategies, random walks, grid based strategies, and brute force exhaustive searching methods[5]. This paper deals with fusion of spatial cluster ensembles using a guided approach to reduce the space complexity of such fusion algorithms.

Spatial data is about instances located in a physical space. Spatial clustering aims to group similar objects into the same group considering spatial attributes of the object. The existing spatial clustering algorithms in literature focus exclusively either on the spatial distances or minimizing the distance of object attributes pairs. i.e., the locations are considered as another attribute or the non-spatial attribute distances are ignored. Much activity in spatial clustering focuses on clustering objects based on the location nearness to each other [5]. Finding clusters in spatial data is an active research area, and the current non-spatial clustering algorithms are applied to spatial domain, with recent application and results reported on the effectiveness and scalability of algorithms [8, 16]. Partitioning algorithms are best suited to such problems where minimization of a distance function is required and a common measure used in such algorithms is the Euclidian distance. Recently new set of spatial clustering algorithms has been proposed, which represents faster method to find clusters with overlapping densities. DBSCAN, GDBDCAN and DBRS are density-based spatial clustering algorithms, but they each perform best only on particular types of datasets [17].

However, these algorithms also ignore the non-spatial attribute participation and require user defined parameters. For large-scale spatial databases, the current density based cluster algorithms can be found to be expensive as they require large volume of memory support due to its operations over the entire database. Another disadvantage is the input parameters required by these algorithms are based on experimental evaluations. There is a large interest in addressing the automation of the general purpose clustering approach without user intervention. However, it is difficult to expect accurate results from the results of these algorithms as each one has its own shortfalls.

B. Litereature on Clustering Ensembles

Clustering ensemble is the method to combine several runs of different clustering algorithms to get an optimal partition of the original dataset. Given dataset $X = \{x_1\ x_2,..\ ,x_n\}$, a cluster ensemble is a set of clustering solutions, represented as $P = P_1,P_2,..Pr$,where r is the ensemble size, i.e. the number of clusterings in the ensemble. Clustering-Ensemble Approach first gets the result of M clusterers, then sets up a common understanding function to fuse each vector and get the labeled vector in the end. The goal of cluster ensemble is to combine the clustering results of multiple clustering algorithms to obtain better quality and robust clustering results. Even though many clustering algorithms have been developed, not much work is done in cluster ensemble in data mining and machine learning community.

Strethl and Ghosh [13,14], proposed a hypergraph-partitioned approach to combine different clustering results by treating each cluster in an individual clustering algorithm as a hyper edge. All the three proposed algorithms approach the problem by first transforming the set of clusterings into a hypergraph representation. Cluster-based Similarity Partitioning Algorithm (CSPA) uses relationship between objects in the same cluster for establishing a measure of pair wise similarity. In Hyper Graph Partitioning Algorithm (HGPA) the maximum mutual information objective is approximated with a constrained minimum cut objective. In their Meta-CLustering Algorithm (MCLA), the objective of integration is viewed as a cluster correspondence problem.

Kai Kang, Hua-Xiang Zhang, Ying Fan [6] formulated the process of cooperation between component clusterers, and proposed a novel cluster ensemble learning technique based on dynamic cooperating (DCEA). The approach is mainly concerned how the component clusterers fully cooperate in the process of training component clusterers.

Fred and Jain [2] used co-association matrix to form the final partition. They applied a hierarchical (single-link) clustering to the co-association matrix. Zeng, Tang, Garcia-Frias and Gao[18], proposed an adaptive meta-clustering approach for

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

combining different clustering results by using a distance matrix.

C. Fusion Framework

We begin our discussion of the guided ensembles fusion framework by presenting our notation. Let us consider a set of n data objects, $D = \{ v_1 \dots v_n \}$. A clustering C of dataset D, is a partition of D into k disjoint sets C1 … Ck. In the sequel we consider m clusterings; we write B_i to denote the i^{th} clustering, and k_i for the number of clusters of B_i. In the clustering fusion problem the task is to find a clustering that maximizes the related items with a number of already-existing clusterings [4].

D. Definations

- Fusion Joint set, FJij:

Fusion Joint set, FJij refers to set of matching pairs of ith clusterer's jth cluster. For instance, FJ12 refers to probable fusion spot for first clusterer's clusters with second clusterer's cluster. It will be used for deciding where the fusion is most likely to yield optimal preciseness of clusters.

- Clusterer Compatibility matrix: CCM (m X m)

Clusterer Compatibility matrix is a m X m symmetric matrix where m is the total number of clusterers, considered for fusion.

- *CCM[i][j]*

Integer value d representing the maximum information gained through the summation of intersection elements cardinality of the matching pairs of clusterer found in Fusion Joint Set, FJ[i][j].

- *Degree Of Agreement Factor: (DOA)*

Degree of agreement factor is the ratio of the index of the merging level to the total number of clusterers. And also this DoA value will be cumulative till it reaches the threshold level DoATh, an user assigned value indicating the majority required for decision making. Under normal scenario, DoATh will be set as 50% of the number of clusterers.

- *Degree of Shadow factor : (DOS)*

Degree of shadow factor is the maximized value of the intersection of the two minimum bounding circles of k clusters with ith cluster from a different clustering.

III. KNOWLEDGE GUIDED ENSEMBLE FUSION

In this section we discuss our proposed layered cluster ensemble fusion guided by the gained knowledge during the merging process. The first phase of the algorithm is the preparation of B heterogeneous ensembles. This is done by executing different clustering algorithms against the same spatial data set to generate partitioning results. For our experimental purpose, we have also generated homogenous ensembles by partitioning the spatial data horizontally/ vertically into n subgroups and used it as the input to our ensemble algorithm. Either way individual partitions in each ensemble are sequentially generated.

A. Selection of clusterings for prime fusion

D Any layered approach will have a drawback of being dependent on which clusterer is considered for initial fusion. This sensitiveness is a major bottleneck in deciding the accuracy of the outputs. But, in our approach, we compute a m x m symmetric clusterer compatibility matrix, where CCM[i][j] indicates the summary of information gain when i^{th} clusterer and j^{th} clusterer are merged. This way we have used heuristics to direct the fusion in the right direction.

B. Resolution for Label Correspondence Problem

The other issues in fusion of cluster ensembles are label correspondence problem and the merging technique used for fusion. At the second phase, we address the label correspondence problem. These clustering results are combined in layered pairs, called fusion joints set, FJ_{mk} . The criteria of merging can be any one of the Fusion Joint Identification Techniques i.e., overshadowing or usage of highest cardinality in intersection set along with usage of add-on knowledge gathered from such association.

First approach uses the degree of shadow that one cluster has on other. This is computed using the smallest circle or minimum covering circle approach, which is a mathematical problem of computing the smallest circle that contains all of a given set of points in the Euclidean plane. Each cluster of the clusterer in two layers first compute the minimum bounding circle and the diameter of such circle, using which the degree of Shadow (DOS) is computed. The aim is to find the clusters in different layers whose shadow overlap is maximized and then assign it to the matching pair set. This method finds the most appropriate clusters belonging to a two clustererings for forthcoming fusion phase.

Second approach uses the usage of heuristic greedy approach in computing mutual information theory to decide on the degree of compatibility. Mutual information is used when we need to decide, which amongst candidate variables are closest to a particular variable. Higher the mutual information, more the two variables are 'closer'. It is the amount of information 'contained' in Y about X.

Let X and Y be the random variables described by the cluster labeling $\lambda^{(a)}$ and $\lambda^{(b)}$, with $k^{(a)}$ and $k^{(b)}$ groups respectively. Let I(X; Y) denote the mutual information between X and Y, and H(X) denote the entropy of X, **i.e,** a measure of the uncertainty associated with a X. The chain rule for entropy states that

$$H(X1{:}n)= H(X1)+H(X2|X1)+...+H(Xn|X1{:}n-1) \qquad (1)$$

When X1:n are independent ,identically distributed (i.i.d.), then H(X1:n) = nH(X1). From Eqn 1, we have

$$H(X, Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$$

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

$$H(X) - H(X|Y) = H(Y) - H(Y|X). \quad (2)$$

This difference can be interpreted as the reduction in uncertainty in X after we know Y, or vice versa. It is thus known as the information gain, or more commonly the mutual information between X and Y. Finding the maximum MI for the clusters in the clusterings is a combinatorial optimization problem, requiring an exhaustive search through all possible clusterings with k labels. This is formidable since for n objects and k partitions there are approximately $k^n/k!$ for n >> k. For example, there are 171,798,901 ways to form four groups of 16 objects. Hence, instead of the complete greedy solution, we have incorporated some heuristics so that cluster accuracy will improve amidst the cost savings in terms of space & computations. As we travel the length and breadth of the ensemble space, we try to reduce the $k^n/k!$ Combinations, by reuse of the cumulative information gain. This way, when n >> k, as in most of the cases of spatial data, the solutions can be reached much faster.

C. Knowledge Guided fusion of clusterings – An Excerpt

Input:

D – the input data in 2-dimensional feature space

Layer : Group of Clusterers B_1 to B_m ;
Levels : List of clusters k_1 to k_n
CCM[i][j] : Clusterer compatability for ith clusterer with jth clusterer.
Step1: Form B_1k_1 to B_ik_n clusters from D using B_1 ~~to~~ B_m clusterers, each clusterer generating k clusters
Step 2; Compute Clusterer compatability matrix, whose entries are the aggregated cardinality values of the intersecting elements set of the clusterers.
Step 3: Identify and select the harmonizing clusterers for fusion from the CCM matrix. as TobeMerged_Layers
Step 4: Set DOA_Increment Factor as 1/ m .
Step 5: Find fusion joints for TobeMerged_Layers , ($FJ_{12\,1}$ .. $FJ_{1\,k}$), using degree of Shadow overlap or maximizing the information gain of probable merge.
Step 6: For every pair in the fusion joint Set, $FJ_{i\,k}$,
Do{
ClustData[i] ← Union of Data points of the pair
Initialize Vector_DOA with DOA_Increment Factor
Append it to to Vector_CData [i]
For each element in the intersection set between Pairs,
DOA[i] ← DOA[i] + DOA_Increment Factor
Increment the vector index I by 1 & merge_ layer by 2
} until i <= k; //normal merge for m/2 layers
Step 7: repeat steps2 to 6 till merge_ layer < m/2;
// finalize the cluster elements at layer i and at level k
do{ If (Vector_DOA > DOA_{Th})
Strong links← Corresponding Elements of Vector_CData
Else
Weak $links_k$ ← Corresponding Elements of Vector_CData
} until all pairs at layer I is resolved
Step 8: Using Strong links, finalize $Final\_Kluster_k$ & continue gathering votes for weak links.
Step 9: From $(m/2+1)^{th}$ layer, perform the pruned merging, where the strong links will be pruned for the confirmed data points, when they reappear. Data points in the weak links could be the noise data points, ($Noise\_Elements_k$), as their inherent votes were below the threshold value.
Step10: Return the robust clusters obtained from m clusterers $Final\_Kluster_k$ and $Noise\_Elements_k$.

Figure 3.3. Excerpt of the guided fusion of ensembles

The initial phase of the fusion starts with finding B_m Clusterers, using m different clustering algorithms. Next stage is to find the clusterers amongst B_m , with maximum compatibility matrix index, for merging, so that they yield maximum knowledge for further fusions. When the merging happens, based on the Fusion Joint Set, for each merged data point, the degree of agreement (DOA) is calculated.. For example, if the total number of clusterers are 5, then all the data points that get merged at level 1, will have DOA as 1/5 = 0.2. This DOA value will be treated as the increment factor for every future fusion. And also the DOA value in the corresponding DOA vector be cumulative till it reaches the threshold level $DOA_{Th.}$ Once the DOA of any point in the cluster crosses the threshold, it can be affirmed to belong to a particular cluster result and will be treated as a strong link. . Thus, the normal voting procedure with huge voting matrix, to confirm the majority does not arise at all in our method.
This final layer merge with the earlier combined clusters will yield the robust combined result. This approach is not computationally intensive, as we tend to use the first law of geography in merging layer by layer. And also the computation of voting matrix is avoided.

The three levels of the technique: fusion joint identification, guided fusion and resolving low voted data points are all executed sequentially. They do not interfere with each other, but they just receive the results from the previous levels. No feedback process happens, and the algorithm terminates after the completion of all procedures.

IV. EXPERIMENTAL PLATFORM & RESULTS

A. The Test Platform

Ensemble Creation : In order to predigest the analysis, the paper uses five representative clustering methods to produce five heterogeneous ensembles or clustering members, viz. DBSCAN, k-means, PAM, Fuzzy K Means and Ward's algorithm. K-means is a very simple yet very powerful and efficient iterative technique to partition a large data set into k disjoint clusters. The objective criterion used in the algorithm is typically the squared-error function. DBSCAN method performs well with attribute data and with spatial data. Partitioning around medoids (PAM) is mostly preferred for its scalability and hence useful in Spatial data. The latest in clustering is the usage of fuzziness and we have added Fuzzy C means (FCM) as one of the clusterer, so that we get a robust partition in the end result. Hence these clustering techniques along with different cluster sizes form the input for our knowledge guided fusion technique.

Data Source **:** Most of the ensemble methods, have sampling techniques in selecting the data for their experimental platform, but this heuristics results in losing some inherent data clusters, thereby reducing the quality of clusters. We have tried to avoid sampling and involve the whole dataset. For our experiments we have used the datasets available in the data repository of University of California, Irvine.

Metrics: We used the classification accuracy (CA) to measure the accuracy of an ensemble as the agreement between the

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

ensemble partition and the "true" partition. The classification accuracy is commonly used for evaluating clustering results. To guarantee the best re-labeling of the clusters, the proportion of correctly labeled objects in the data set is calculated as CA for the partition. We have used the measurement of intra cluster and inter cluster density before and after usage of our cluster ensemble approach, which will be a metric for the preciseness of the so formed cluster groups.

B. *Validation of fusion Results*

As clustering is a kind of study without guidance, basically unsupervised classification, it is difficult to evaluate the clustering efficiency. But with classifying information of data, it can be considered that some inner distribution characters of the data are expressed to certain degree. If such classifying information is not used by clustering, it can be used to evaluate the clustering effect. If the number of same objects, which covered by certain clustering label of labeled vectors and certain known category of category properties, are at best, this clustering label corresponds to this known category. Thus many clustering labels might correspond to the same category, whereas one clustering label can not correspond to many categories. The clustering results can be evaluated by classifying information.

The test results with the IRIS dataset, Wine dataset, Half rings and Spiral dataset (Courtesy: UCI data repository) is promising and shows better cluster accuracy. Two parameters were computed to verify our algorithm: Cluster Correctness Factor (CCF) and the space complexity of the fusion of ensembles. Few bench marked datasets as mentioned above were tested with this technique and the CCF was found to be 100%, in all the cases. Normally, in all ensembling algorithms, voting matrix is computed which is normally in the order of n^2, where n is the number of data points. But, due to the knowledge guided fusion along with unique inherent voting scheme, the space complexity has been reduced to the order of n. This has a major impact in not only memory requirements but also in the total number of matrix computations.

C. *Comparison of the Experimental Results*

In our approach of knowledge guided fusion, we have combined the results of several independent cluster runs by computing inherent voting of their results. Our phased knowledge guided approach voting helps us to overcome the computationally infeasible simultaneous combination of all partitions and also increases the cluster accuracy. (Figure 4.3.1). By the help of our scheme, we have shown that the consensus partition indeed converges to the true partition. InterCluster Density (Figure 4.3.3) has been reduced by almost 40% when compared against the other clustering algorithms with our technique. For the benchmark iris dataset it is around 11.47 and our cluster miner produces 6.77 implying that the later has produced better cluster in terms intercluster Density. We have observed that the IntraCluster Density (Figure 4.3.2) has increased, implying that the cluster quality has improved due to the guided approach used for ensemble fusion. For the standard benchmark iris dataset, intra cluster density achieved using normal clustering methods is 5.1283, whereas our technique had generated an intra cluster density of 5.7125, implying that we have generated more precise clusters.

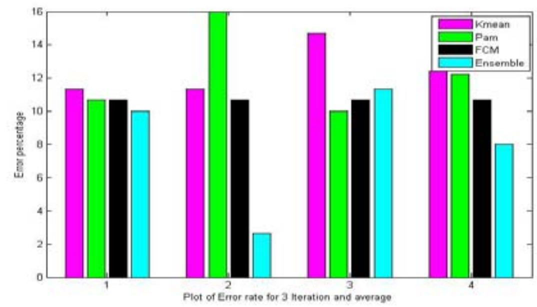


Figure 4.3.1 Comparison of error rates of the fused ensembles

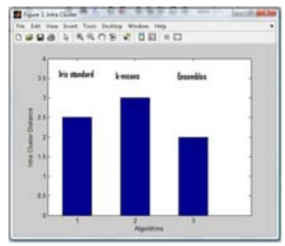


Figure 4.3.2 Intra cluster density

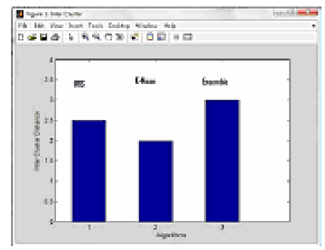


Figure 4.3.3 Inter cluster density

With Iris dataset

V. CONCLUSION AND FUTURE WORK

In this paper we addressed the relabeling problem found in general in most of cluster ensembles, and has been resolved without much computations, using the notion from first law of geography. The cluster ensemble is a very general framework that enables a wide range of applications. We have applied the proposed guided cluster merging technique on spatial databases. The main issue in spatial databases is the cardinality of data points and also the increased dimensions. Most of the existing Ensemble algorithms have to generate voting matrix of at least an order of $n^{2.}$ or an expensive graphical representation with the vertices which is equal to the number of data points. When n is very huge and is also a common factor in spatial datasets, this restriction is a very big bottleneck in obtaining robust clusters in reasonable time and high accuracy.

Our algorithm has resolved the re labeling using layered merging as well as guided by the gained information. Once elements move from strong links to final clusters, they do not participate in further computations. Hence, the computational cost is also hugely reduced. Usage of the Cluster compatibility matrix enables us to have a good head start in the fusion process, which otherwise is a matter of sheer randomness.

The key goal of *spatial* data mining is to automate knowledge discovery process. It is important to note that in this study, it has been assumed that, the user has a good knowledge of data and of the hierarchies used in the mining process. The crucial input of deciding the value of k, still affects the quality of the resultant clusters. Domain specific Apiori knowledge can be

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

used as guidance for deciding the value k. We feel that semi supervised clustering using the domain knowledge could improve the quality of the mined clusters. We have used heterogeneous clusterers for our testing but it can be tested with more new combinations of spatial clustering algorithms as base clusterers. This will ensure exploring more natural clusters.

First, we have identified several non-spatial datasets which are normally used as bench mark ones for data clustering. Then we tested how our layer based methodology can wok with spatial data. This setup must be worked with more large datasets available in GIS areas and with satellite images. We evaluated our work and can conclude that for targeting a specific platform and incorporating spatial feature space, our automated layered merge approach is able to provide the necessary correctness with more efficiency both in space constraint and in matrix computations. However, more work should be carried out to provide support for more real life data from satellites and incomplete data. Future work in the short term will focus on how to acquire such datasets, and continue with more testing, in spite of current security concerns in distributing such data.

ACKNOWLEDGMENT

This work has been partly done in the labs of The Oxford College of Engineering, Bangalore, where the author is currently working as a Professor, in the department of Computer Science & Engineering. The authors would like to express their sincere gratitude to the Management and Principal of The Oxford College of Engineering for their support rendered during the testing of some of our modules. They also express their thanks to the University of California Irvine, for their huge data repository made available for testing our knowledge guided approach of fusion of ensembles.

REFERENCES

[1] M.Ester, H. Kriegel, J. Sander, X. Xu. "Clustering for Mining in Large Spatial Databases". Special Issue on Data Mining, KI-Journal Tech Publishing, Vol.1, 98.

[2] A.L.N. Fred and A.K. Jain, "Data Clustering using Evidence Accumulation". In Proc. of the 16th International Conference on Pattern Recognition, ICPR 2002, Quebec City.

[3] A.L.N. Fred and A.K. Jain, "Robust data clustering" in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR, USA, 2003.

[4] Filkov, V. and Skiena, S. "Integrating microarray data by concensus clustering". In International Conference on Tools with Artificial Intelligence, 2003

[5] K.Koperski, J.Han, K. Koperski and J. Han, "Discovery of spatial Rules in Geographic Information Databases," Proc. 4th Intl Symposium on Large Spatial Databases, pp. 47-66, 95.

[6] Kai Kang, Hua-Xiang Zhang, Ying Fan, "A Novel Clusterer Ensemble Algorithm Based on Dynamic Cooperation", IEEE 5TH International Conf. on Fuzzy Systems and Knowledge Discovery 2008.

[7] Matheus C.J., Chan P.K, and Piatetsky-Shapiro G, "Systems for Knowledge Discovery in Databases", IEEE Transactions on Knowledge and Data Engineering 5(6), pp. 903-913, 1993.

[8] Ng R.T., and Han J., "Efficient and Effective Clustering Methods for Spatial Data Mining", Proc. 20th Int. Conf. on Very Large DataBases, 144-155, Santiago, Chile, 1994.

[9] B.H. Park and H. Kargupta, "Distributed Data Mining", In The Handbook of Data Mining, Ed. Nong Ye, Lawrence Erlbaum Associates, 2003.

[10] J. Roddick and B. G. Lees, "Paradigms for Spatial and Spatio-Temporal Data Mining," in Geographic Data Mining and Knowledge Discovery, Taylor & Francis, 2001.

[11] Su-lan Zhai1,Bin Luo1 Yu-tang Guo : Fuzzy Clustering Ensemble Based on Dual Boosting , Fourth International Conference on Fuzzy Systems and Knowledge Discovery 07.

[12] Samet, Hanan.: "Spatial Data Models and Query Processing". In Modern Databases Systems: The object model, Interoperability, and Beyond. Addison Wesley/ ACM Press, 1994,Reading, MA.

[13] A.Strehl, J.Ghosh, "Cluster ensembles - a knowledge reuse framework for combining multiple partitions", Journal of Machine Learning Research, 3: 583-618, 2002.

[14] A.Strehl, J.Ghosh, "Cluster ensembles- a knowledge reuse framework for combining partitionings", in: Proc. Of 11th National Conference On Artificial Intelligence, NCAI, Edmonton, Alberta, Canada, pp.93-98, 2002.

[15] Y. Tao, J. Zhang, D. Papa dias, and N. Mamoulis, "An Efficient Cost Model for Optimization of Nearest Neighbor Search in Low and Medium Dimensional Spaces," IEEE Transactions on Knowledge and Data Engineering, vol. 16,no. 10, pp. 1169-1184, 2004.

[16] X. Wang and H. J. Hamilton, "Clustering Spatial Data in the Presence of Obstacles," International Journal on Artificial Intelligence Tools, vol. 14, no. 1-2, pp. 177-198, 2005.

[17] R. Xu and D. Wunsch, II, "Survey of clustering algorithms," IEEE Transactions on Neural Networks, vol. 16,no. 3, pp. 645- 678, 2005.

[18] Zhang, J. 2004. Polygon-based Spatial clustering and its application in watershed study. MS Thesis, University of Nebraska-Lincoln, December 2004.

[19] Zeng, Y., Tang, J., Garcia-Frias, J. and Gao, G.R., "An Adaptive Meta-Clustering Approach: Combining The Information From Different Clustering Results", CSB2002 IEEE Computer Society Bioinformatics Conference Proceeding.

AUTHORS PROFILE

RJ Anandhi is a PhD student in the department of Computer Science & Engineering at Dr M G R University She is currently working as a professor in the Department of Computer Science at Oxford College of Engineering, Bangalore. She has completed her BE degree from Bharatiyar University and MTech degree from Pondicherry Central University. Her research interests are in Spatial Data mining and ANT algorithms.

Dr. Natarajan has initially worked in Defence Research and Development Laboratory (DRDL) for five years in the area of software development in defence missions. Dr Natarajan then worked for 28 years in National Remote Sensing Agency (NRSA) in the areas pertaining to DIP and GIS for several remote sensing missions like IRS-1A, IRS-1B, IRS-1C, IKONOS and LANDSAT. As a Project Manager of Ground Control Point Library (GCPL) Project, he had completed the task of computing cm level accuracy for 3000 locations within India which is being used for cartographic satellite missions. He was the Deputy Project Director of Large Scale Mapping (LSM) of Department of Space. Dr Natarajan has published about fifteen papers in National/ International Conferences and Journals. His research interests are Data mining, GIS and Spatial Databases.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Fault Diagnosis Algorithm for Analog Electronic Circuits based on Node-Frequency Approach

S.P. Venu Madhava Rao

madhavaraosp@gmail.com

Dr. N. Sarat Chandra Babu & Dr. K. Lal Kishore

Abstract: In this paper we present a novel approach to analog electronic circuits fault diagnosis based on selection of both nodes and frequency for the first time as far as we know. Two fault isolation and localization algorithms are presented in this paper. The first algorithm selects nodes and frequencies which isolate all or desired number of faults. The second algorithm presented converts the fault dictionary contents into binary form. Importantly this helps in the automation of the fault diagnosis process.

Keywords: Fault Dictionary, Fault Isolation Table, Binary dictionary, singletons.

I. Introduction

Analog Fault Diagnosis has been of immense research interest for the past three decades and continues to sustain the same zeal even today. The main challenges today in analog fault diagnosis are to design universally accepted fault models, cost effective, faster and accurate diagnosis of faults. Importantly all this is desired even in the presence of inherent characteristics of analog circuits like tolerances, non linearity, in accessible test nodes etc.

There are two categories of analog circuit fault diagnosis: Simulation before test (SBT) and Simulation after test (SAT) [1]. The SBT approach involves the generation of fault dictionary by simulating the circuit and then using pattern recognition to identify the faults. This is the most popular method adopted. In SAT approach sufficient measurements are needed to identify faulty parameters. In the SBT approach construction of fault dictionary is an efficient method. Different test measurements like node voltages, current sources, branch currents, frequency measurements etc are used in the construction of fault dictionaries [2]. There are some algorithms developed to find out testable measurements using numerical approach in [3] and [4].

In [5] a new method in the construction of fault dictionary is proposed where a combination of sensitivity based and information channel based approaches are used. Also the construction of integer coded fault dictionary using Quasi-Hamming distance is proposed in this paper. Heuristic methods using evolutionary computation in combination with the Fuzzy logic is presented in [6], the main purpose of such a combination is to generate an optimized frequency test set and also ambiguity sets are provided to avoid take care of tolerance effects. An SBT based approach is proposed in [7] where the fault dictionary is constructed using test node voltages and the method used to approximate is Section wise piecewise linear (SPLF) method. A procedure for the selection of test frequencies is presented in [8]. This is based on the evaluation of algebraic indices and the inverse norm of a sensitivity matrix of the circuit under test. In [9], [10] and [11], fault diagnosis based on different types of neural networks has been proposed. In [12] knowledge base and fuzzy logic have been used in fault diagnosis. The knowledge base is developed in two ways, one by simulations and the second is based on heuristic symptoms observed by the operator. In [13] the ambiguity sets are divided based on the lowest error probability in the construction of fault dictionaries is proposed. This paper used Monte Carlo techniques for sensitivity analysis. In [14] a fault threshold function and a fault criterion have been proposed for the fault diagnosis of circuits with tolerance. An algorithm is proposed in [15], which aims to reduce the size of the fault dictionary. In [16] and [17] different methods and algorithms are used to reduce the size of the fault dictionaries. In [18] time slot specification based approach is used in analog fault diagnosis. For this built in current sensors and test point insertion is used. A sensitivity based approach using randomized algorithms is used to diagnose soft faults in [19].In [20] the algorithm proposed tries to find the minimum number of test point for maximum fault isolation. This approach is based on information measure of the test

 http://sites.google.com/site/ijcsis/
ISSN 1947-5500

points. The diagnosis proposed in this paper [21] is based on global sensitivity analysis method. Also fuzzy logic is used to obtain the sensitivity curves. In [22] an efficient method is applied in the selection of test nodes. This is done by searching for the minimum entropy index based on the available test points. An efficient graph based method is proposed in [23].This method can be used to select optimum test point selection and also can be sued to build DFT. Efficient Inclusion methods and Exclusion methods are proposed in [24] to select or de select test nodes, in other words the faster selection of optimum test points. A novel multi frequency approach is proposed in [25] which drastically reduce the number of test frequencies needed to achieve maximum fault diagnosis. The reduction achieved is better than any known methods. The method proposed in [26] consists of two parts. One is the creation of fault dictionary consisting of nominal and faulty states of the components and second is a novel fault detection and localization algorithm.

This paper proposes a novel approach where both test node and multi frequency techniques are used. This approach is used to diagnose all the faults or the desired number of faults.

II Node-Frequency Approach

In the analog fault diagnosis the prominent methods used are multi node or multi frequency measurements. The research so far has been on developing methods to find out optimum number of test nodes or test frequencies that can identify the desired faults. This in some cases leads to more number of measurements being made thus drastically increasing the size of the dictionary.

In this paper we have taken basically nodal analysis and then a choice of test frequencies is made based on [27]. The proposed algorithm selects the nodes and frequencies which isolate all or desired faults.

In this paper two algorithms are presented. The first algorithm is for fault isolation and localization. The second algorithm converts the integer coded fault dictionary into a binary dictionary which helps in faster fault isolation.

The actual measurements of the CUT are noted down and these values are normalized if necessary. From these values we form ambiguity sets. Now we construct another table called integer coded table using ambiguity sets. Then the original readings are replaced by integer numbers indicative of the ambiguity set to which these values belong.

The test frequency set is represented by f_I to f_M, where N is the number of frequencies chosen.

The nodes are represented by n_I to n_P, where P represents the total number of nodes.

The faults are represented by F_0(nominal value) to F_N, where N represents the total number of faults.

Algorithm 1:

Step 1: Select the test frequency set (f_I to f_M).

Step 2: Select the test nodes (n_I to n_P) which are accessible for each frequency.

Step 3: Note the actual readings of the circuit for the test frequency set and nodes chosen in steps 1 and 2.

Step 4: Form the integer coded dictionary using the ambiguity sets.

Step 5: Identify unique integer codes called singletons for each row i.e. for each of the nodes selected.

Step 6: Identify the node (n_K) which has maximum number of singletons for a frequency f_J, where $1<K\leq P$ and $1<J\leq M$. Select this node-frequency (n_K, f_J) pair. If more than one node satisfies this condition, then go to step 9.

Step 7: If the number of singletons is equal to N+1, then go to step 12. If else go to next step 8.

Step 8: Call Algorithm 2, to form binary dictionary which helps in identifying other nodes from the remaining (P-1) nodes belonging to the frequency f_J, which can identify different faults. If all faults are isolated then go to step 12.

Step 9: Find the total number of singletons for each test frequency. Then choose the node belonging to the frequency which has the maximum number of singletons. If more than one frequency satisfies this condition choose any one of the nodes randomly.

Step 11: If all the faults or desired number of faults are not isolated, then repeat steps from 6 with the next highest number of singletons.

Step 12: Stop

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Algorithm 2:

Step 1: Replace all the singletons by the value '1' and others by '0' in the integer coded table, resulting in a binary table.

Step 2: If n_K is the node chosen, then calculate (n_M-n_K), where $1<M\leq P$, thus forming another table called Node-Wise Fault Isolation table. This results in three values 0,-1 or 1. The value '0' indicates that the fault has been identified by both n_M and n_K or both the nodes did not isolate the fault, whereas '-1' indicates that the fault has been isolated by only n_K and '1' is an indication that the fault has been identified by the node of interest i.e. n_M. Therefore choose the node n_M which has maximum number of 1's.

Step 3: Check the total number of faults isolated by the nodes n_k and n_M. If this sum is equal to P, then Stop, otherwise choose the node which has the next highest number of 1's.

Step 4: Repeat step 3 till the desired fault isolation is achieved or no further isolation is possible.

Step 5: Return to Algorithm 1

III. Integer coded dictionary based on ambiguity sets

The formation of the Integer coded dictionary based on ambiguity sets is illustrated by an example in this section. Assume that the actual readings of an imaginary circuit under test are given in Table 1 below.

Table 1: Actual readings of the imaginary CUT

Node	Nominal	Fault-1	Fault-2	Fault-3
Node-1	1.22	0.33	0.78	0.34
Node-2	1.33	0	0.09	2.1
Node-3	1.45	0	0.99	2.5

As seen from the Table 1 above, we see that for node -1 measurement, fault-1 and fault -3 have almost the same value and thus belong to the same ambiguity group. Also these two values are the least among all and are assigned values '1'. The other values do not belong to any ambiguity group and are assigned values 2 for fault-2 and 3 for nominal, based on the ascending range of the values. Using the same procedure for all the remaining nodes, integer coded fault dictionary is formed and is shown in Table 2.

Table 2: Integer Coded Fault Dictionary

Node	Nominal	Fault-1	Fault-2	Fault-3
Node-1	3	1	2	1
Node-2	2	1	1	3
Node-3	3	1	2	4

In the Table 2, we see that node-1 has 2 singletons, node-2 has 2 singletons and node 3 has 4 singletons.

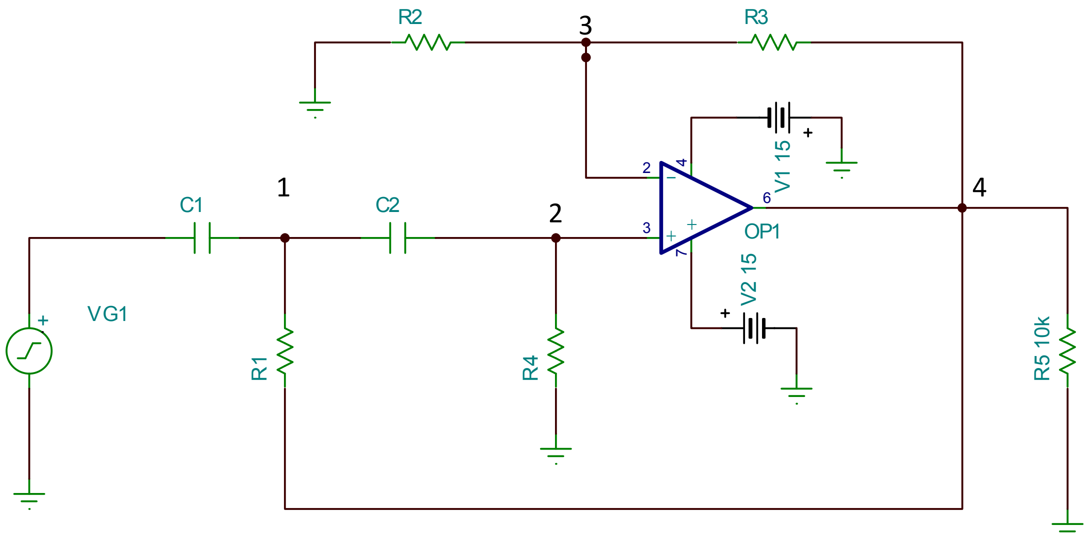
IV Illustration

The circuit used here is a 2nd order Butterworth High Pass Filter as shown in Fig. 1. The circuit has been simulated using Tina Spice software.

The faults chosen are taken as 50% increase or decrease in the component values. Thus CUT has been simulated for these faults by changing the component values by ±50%.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Figure 1: Second Order Butterworth-High Pass Filter



Using the Step1 from the Algorithm 1, we have chosen the test frequency f_T= {500Hz, 800 Hz, 1000Hz, 1200Hz, and 1500Hz}. From Step 2, we have chosen four nodes with the assumption that all these nodes are accessible. Using Step3 and 4, the CUT has been simulated and the integer coded dictionary as shown in Table 3 is formed based on the actual readings.

Table 3: Integer coded Dictionary for the HP Filter

Frequency=500Hz													
Nodes/Faults	F0	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12
N1	***5***	7	***2***	4	7	6	4	8	***3***	8	***1***	6	4
N2	2	3	1	2	3	2	2	***5***	***6***	4	1	4	1
N3	2	3	1	2	3	2	2	***5***	***6***	4	1	4	1
N4	***5***	***6***	***2***	***4***	***9***	7	***3***	***10***	1	7	1	***8***	1
Frequency =800 Hz													
N1	***7***	***6***	***3***	***5***	***11***	***10***	4	***12***	***2***	***9***	***1***	***8***	4
N2	***6***	5	***4***	5	***9***	7	5	***10***	***1***	7	***3***	***8***	***2***
N3	***6***	5	***4***	5	***9***	7	5	***10***	***1***	7	***3***	***8***	***2***

N4	***7***	***6***	5	5	***12***	***10***	***4***	***11***	***1***	***8***	***3***	***9***	***2***
Frequency=1000Hz													
N1	***4***	3	5	3	***8***	***6***	3	***7***	***1***	5	***2***	5	3
N2	***6***	5	7	5	***9***	8	***4***	***10***	***1***	7	***3***	8	***2***
N3	***6***	5	7	***5***	***9***	8	***4***	***10***	***1***	7	***3***	8	***2***
N4	***6***	***5***	***7***	***4***	***12***	***10***	3	***11***	***1***	***8***	3	***9***	***2***
Frequency= 1200 Hz													
N_1	***5***	3	***9***	4	***10***	***7***	3	***8***	***1***	6	***2***	6	4
N_2	6	4	***9***	***5***	***11***	***8***	4	***10***	***1***	6	***3***	***7***	***2***
N_3	6	4	***9***	***5***	***11***	***8***	4	***10***	***1***	6	***3***	***7***	***2***
N_4	***6***	***5***	9	4	***11***	9	***3***	***10***	***1***	***7***	4	***8***	***2***
Frequency=1500Hz													
N_1	5	2	***8***	3	***7***	6	2	6	***1***	5	3	5	***4***
N_2	4	3	***9***	3	***8***	***6***	3	***7***	***1***	4	3	***5***	***2***
N_3	4	3	***9***	3	***8***	***6***	3	***7***	***1***	4	3	***5***	***2***
N_4	7	***5***	11	***4***	11	***10***	***3***	***9***	***1***	7	***6***	***8***	***2***

The number of singletons for each node for the whole frequency set is calculated (Step 5) and tabulated in Table 4. Here the frequencies are f_1=500Hz, f_2=800Hz, f_3=1000Hz, f_4=1200Hz and f_5=1500Hz.

Table 4: Total number of singletons

Node/Freq	f_1	f_2	f_3	f_4	f_5
N_1	4	11	6	7	4
N_2	2	8	7	9	7
N_3	2	8	7	9	7
N_4	8	11	11	9	9
Total	**16**	**38**	**31**	**34**	**27**

As seen from the Table 4, node1 and node 4 of frequency set f_2, and node 4 of frequency set f_3 have maximum number of singletons equal to 11, i.e. these nodes can isolate 11 of the total thirteen faults. We have chosen node 1(or even node 4 can be chosen) of the frequency set f_2 i.e. 800Hz as it has maximum number of total singletons (step 9). As the condition mentioned in step 7 is not satisfied, binary table is formed as per step 8.

The binary table is formed replacing Table 3 contents by either '0' or '1'. All the singletons are replaced by '1'and ambiguity sets by '0' (step 1 of Algorithm 2). The binary fault dictionary is shown in Table 5.

After the execution of the step 2(Algorithm 2), the results are shown in the Node-wise Fault isolation Table 6.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Table 5: Binary Dictionary

Frequency: 800Hz													
Nodes	F_0	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}	F_{11}	F_{12}
N_1	1	1	1	1	1	1	0	1	1	1	1	1	0
N_2	1	0	1	0	1	0	0	1	1	0	1	1	1
N_3	1	0	1	0	1	0	0	1	1	0	1	1	1
N_4	1	1	0	0	1	1	1	1	1	1	1	1	1

Table 6: Node-Wise Fault Isolation Table

Frequency: 800Hz													
Nodes	F_0	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}	F_{11}	F_{12}
N_1	1	1	1	1	1	1	0	1	1	1	1	1	0
N_2-N_1	0	-1	0	-1	0	-1	0	0	0	-1	0	0	1
N_3- N_1	0	-1	0	-1	0	-1	0	0	0	-1	0	0	1
N_4- N_1	0	0	-1	-1	0	0	1	0	0	0	0	0	1

From the Binary dictionary of Table 5, we can see that the faults isolated are F_0, F_1, F_2, F_3, F_4, F_5, F_7, F_8, F_9, F_{10} and F_{11}, where as faults F_6 and F_{12} are not isolated. The faults not isolated is deduced from the '0' entry in the corresponding columns. As seen from the Table 6, the total number of 1's is two for (N_4-N_1), one for (N_3-N_1) and (N_2-N_1). So we choose the (N_4-N_1) column, i.e. node 4 is chosen. The faults isolated by this node 4 are F_6 and F_{12}. As seen these are the faults which are not isolated by node 1.

In the example discussed in this paper we have been able to achieve 100% fault diagnosis. This is achieved by a single test frequency of 800Hz and nodes 1 and 4.

V. Conclusions

In this paper we have presented a novel method using node-frequency approach in analog fault diagnosis. We have presented two algorithms, the first one for choosing the frequencies and nodes for the desired fault isolation and the second is for the generation of binary dictionaries. The effectiveness of these two algorithms was demonstrated using a HP filter circuit.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

References

[1] Tsung-Chih Lin, "Analog circuit fault diagnosis under parameter variations based on Type-2 Fuzzy logic systems", International Journal of Innovative Computing, Information and Control, Vol. 6, No. 5, pp. 2137, May 2010.

[2] Jansuz A. Starzyk, Jing Pang, Stefano Manetti and Maria Cristina Piccirilli and Giulio Fedi, " Finding ambiguity groups in low testability analog circuits", IEEE Transactions on circuits and systems-I: Fundamental theory and applications, vol. 47, NO.8, August 2000.pp.1125-1137.

[3] G.Luculano,A.Liberatore,S.Manetti and M.Marini, "Multi frequency measurement of testability with application to large linear analog systems", IEEE Transactions on Circuits and Systems, Vol. CAS-23, pp. 644-648, June 1986.

[4] M.Catelani, G.Luculano, A.Liberatore, S.Manetti and M.Marini, "Improvements to numerical testability evaluation", IEEE Transactions on Instrumentation and Measurements, Vol. IM-36, pp. 902-907, December 1987.

[5] Jerzy Rutkowski and Jan Machniewski, " Integer code DC fault dictionary", ISCAS 2000- IEEE International Symposium on Circuits and Systems, May 28-31.pp 713-716

[6] P.Jantos, D.Grzechca, T.Golenek and J.Rutkowski, "Heuristic methods to test frequencies optimization for analogue circuit diagnosis", Bulletin of the Polish Academy of Sciences, Technical Sciences, Vol. 56, No. 1, pp. 29-38, 2008.

[7] S.Halgas, "Multiple fault diagnosis of non linear circuits using the fault dictionary approach", Bulletin of the Polish Academy of Sciences, Technical Sciences, Vol. 56, No. 1, 2008.

[8] Franseco Grasso, Antonio Luchetto, Stefano Manetti and Maria Cristina Piccirilli, "A method for the automatic selection of test frequencies in analog fault diagnosis", IEEE Transactions on Instrumentation and Measurement, Vol. 56, No. 6, December 2007

[9] Wei- Qiang Zhang and Chen Xu, "Improved algorithms for circuit fault diagnosis based on wavelet packet and neural network", International Symposium on Non linear dynamics, Journal of Physics: Conference series 96, pp. 1-7, 2008.

[10] Farzan Aminian Mehran Aminian and H.W.Collins, "Analog fault diagnosis of actual circuits using Neural networks", *IEEE Transactions on Instrumentation and Measurements,* Vol. 51, No.3, pp. 544-549, June 2002.

[11] K.Mohammadi, A.R. Mohseni Monfarad and A.Molaei Nejad, "Fault diagnosis of analog circuits with tolerances by using RBF and BP Neural Networks", *Student Conference on Research Development Proceedings, IEEE,* pp. 317-321, 2002.

[12] Lamiaa Mohamed and A.S. Ibrahim, "Model based Fault diagnosis via parameter estimation using knowledge base and Fuzzy logic approach", *IEEE MELECON,* pp. 505-509, May 2002.

[13] Jinyan Cai and M.S.Alam, "An algorithm dividing ambiguity sets for analog fault dictionary", IEEE, 2002.pp89-92

[14] Peng Minfing and He Yigang, "Fault dictionary diagnosis based on branch screening for tolerance circuits", ICSP'04 proceedings, pp. 1488-1491.

[15] P.Bernardi, M.Grosso, M.Rabaudengo and M.Sonza Reorda, "A pattern ordering algorithm fro reducing the size of fault dictionaries", Proceedings of the 24th IEEE VLSI Test Symposium (VTS'06), 2006.

[16] David B.Lavo and Tracy Larrabee, "Making cause-effect cost effective: Low-Resolution fault dictionaries", *ITC*

International Test Conference, IEEE, pp. 278-286, 2001.

[17] Baris Arslan and Alex Orailoglu, "Fault dictionary size reduction through test response superposition", *IEEE International Conference on Computer Design: VLSI in Computers and Processors, IEEE,* 2002.

[18] Shambhu Upadhyaya, Jae Min Lee and Padmanabhan Nair, "Tie slot specification based approach to analog fault diagnosis using built-in current sensors and test point insertion", Proceedings of the 11th Asian Test symposium (ATS'02), 2002.

[19] Cesare Alippi, Marcantonio Catelani, Ada Fort and Marco Mugnaini, "SBT Fault Diagnosis in analog electronic circuits: A sensitivity-based approach by randomized algorithms", *IEEE Transactions on Instrumentation and measurement,* Vol. 51, No. 5, pp. 1116-1124, October 2002.

[20] Kranthi K.Pinjala and Bruce C.Kim, "An approach for selection of test points for analog fault diagnosis", Proceedings of the 18th IEEE International Symposium on Defect and Fault Tolerance in VLSI systems (DFT'03), 2003.

[21] C.Alippi, M.Catelani, A Fort, M.Mugnaini, "Automatic selection of test frequencies for the diagnosis of soft faults in Analog circuits", IEEE Instrumentation and Measurement Technology conference, May 2002. pp 1503-1508

[22] Jansuz A. Starzyk, Dong Liu, Zhi-Hong Liu, Dale E. Nelson and Jerzy O. Rutkowski, "Entropy based optimum test points selection for analog fault dictionary techniques", IEEE Transactions on Instrumentation and Measurement, Vol. 53, No. 3, pp. 754-761, June 2004.

[23] Jiun-Lang Huang and Kwang-Ting Cheng, "Test point selection for analog fault diagnosis of unpowered circuit boards", IEEE transactions on circuits and systems-II: Analog and Digital signal processing, vol.47, No. 10, October 2000. pp 977-987.

[24] V.C.Prasad and N.Sarat Chandra Babu, "Selection of Test nodes for analog fault diagnosis in dictionary approach", IEEE transactions on Instrumentation and measurements, vol. 49, NO.6, pp.1289-1297, December 2000.

[25] N.Sarat Chandra Babu, V.C.Prasad, S.P. Venu Madhava Rao and K.L.Kishore, "Multi-Frequency approach to fault dictionary of linear analog fault diagnosis", Journal of Circuits, Systems and Computers, Vol. 17, No. 5, pp. 905-928, October 2008.

[26] Zbigniew Czaja, "A fault diagnosis algorithm of Analog circuits based on Node-Voltage relation", 12th IMEKO TC1 & TC7 Joint Symposium on Man Science & Measurement, pp. 297-304, 2008.

[27] S.Seshu and R.Waxman, "Fault isolation in conventional linear systems- A feasibility study", IEEE Transactions on Reliability, R-15, pp. 11-16, 1986.

Significance of Rapid Solutions Development to Business Process Management

Steve Kruba

Northrop Grumman
3975 Virginia Mallory Drive, Chantilly VA 20151, USA
steve.kruba@ngc.com

***Abstract*—Business process management (BPM) is moving from a niche market into the mainstream. One of the factors leading to this transformation is the emergence of very powerful rapid solutions development tools for creating BPM solutions (BPM RSD). It has been widely recognized that this facility is important for achieving benefits quickly. Similar benefits are attributed to the agile software movement, but BPM RSD differs in that the objective is to *reduce* the need for custom software development. As the BPM RSD features of some of the current business process management suites (BPMS) products have matured, additional benefits have emerged that fundamentally change the way we approach solutions in this space.**

***Keywords*—BPM, Business process management, workflow, agile, rapid applications development, rapid solutions development, RAD, BPM RSD, BPM RAD.**

I. INTRODUCTION

Technology, in a traditional sense, is not the differentiator that attracts customers. In the BPM context, the "technology" is about how the extensive functionality that is required for successfully automating a customer's mission-critical business processes is applied to solving their problem.

Business users are not information technology (IT) experts and have difficulty relating technical designs to their business needs. Furthermore, most business users have great difficulty articulating their needs since they have little experience or involvement working with complex process solutions. This has historically been a major impediment to creating successful BPM solutions.

Modern BPMS products provide a rich application development infrastructure with significant out-of-the-box capabilities and extensive hooks for customization. This paper will provide information on these capabilities and the benefits that are provided. Not only do these capabilities provide a rich environment for building solutions, but the combination of rapid solutions development and the rich internal constructs needed to support it amplify a designer's ability to conceptualize these solutions. By providing the major, base functionality, these products allow architects and developers to focus on the unique aspects of each solution – the issues that make the difference between successful and unsuccessful projects. Examples will be provided based on Northrop Grumman's e.POWER®[1] BPMS product.

[1]e.POWER is a commercial BPM product and a registered trademark of the Northrop Grumman Corporation.

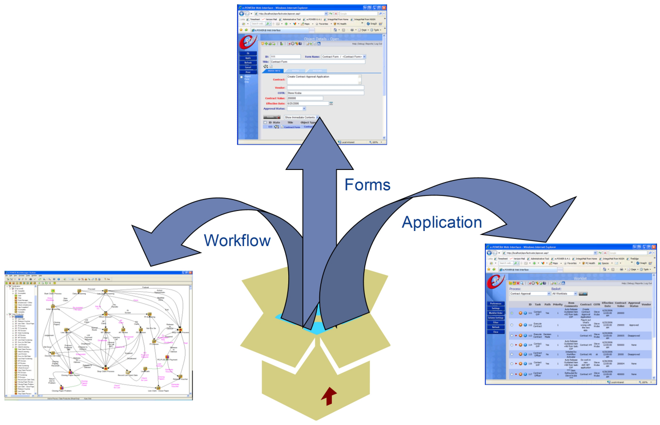


Fig. 1. Three Key BPM Components

There are over 100 products in the BPM software product market as well as products servicing other software product segments that have BPM features. A small subset of these products offer the capabilities described in this paper. The implications of these capabilities are, perhaps, more significant than have previously been documented, and affect all aspects of the system development life cycle (SDLC).

II. BPM RSD FEATURE REQUIREMENTS

BPM RSD tools focus on providing the three key components required for any BPM solution: the business process or workflow, an application for doing the work, and forms as the basis for user interaction. These three components are illustrated in Figure 1. The extent to which a particular BPM product provides these capabilities out-of-the-box is a measure of their "out-of-the-boxness." Keep in mind that not all BPM products have BPM RSD toolsets.

Automating a business process involves two key steps:

❶ Creating an automated representation of the business process – see Figure 2. Drag and drop interfaces are the norm with BPM RSD tools. Note that in addition to being a visual representation of the process, it also defines the rules for the process in a backend store that is later used by the process engine for managing the work. Engines of this type are said to be "model-driven" because changes to the model directly affect production instances of the process. Other, less flexible approaches include configuration-driven and parameterized where a

limited set of options are baked in by the product vendor. [1]

❷ Creating an application for users to process their work (see Figure 3) that is *process-enabled*. For the toolset to be considered a BPM RSD toolset, the user interface should be a byproduct of the application definition process – a declarative process rather than a programming exercise. While it is important to automatically generate the user interface, it is also important to provide customization hooks needed to tweek the interface, since rarely is the one-size-fits-all approach adequate.

III. Agility

The primary purpose of process automation is process improvement. Complex business processes are constantly changing – with or without explicit direction. Factors such as changing business environments, government regulation, and competition are major drivers of these changes, necessitating changes in the support systems.

The traditional life-cycle development approaches to custom development are seriously challenged to support these dynamics.[2] Historically, requirements documentation, detailed systems design, development, and implementation could easily take 12 to 18 months to deliver a complex solution, during which time the business requirements may have changed significantly enough to require additional iterations prior to implementation.

Agility has become a popular term for describing the flexibility needed by organizations to operate in today's dynamic environments. Agility is a natural by-product of BPM RSD toolsets. Agility is a critical feature of BPMS products.

Agility in the BPM context is similar to, but not the same as the agile software development methodology, typically used in an iterative process for creating custom software. Agility in the BPMS space is more about using the built-in capabilities of the BPM product to *avoid* having to write custom software. Custom software is needed as part of the creation process for most BPM solutions, but whenever it can be avoided, the resulting solution is less expensive and has fewer defects and lower risk. In order to differentiate this process from rapid applications development (RAD) approaches, I have coined the term "BPM RSD."

Another related software engineering concept is model-driven development. These techniques often include a framework in which software is developed, providing a powerful facility for leveraging the assets within the framework for reuse.

The concept of 'models' is critical to BPM products where the software architecture creates models of the organization's business operations – a key example being the model of the operational aspects of the business being encapsulated in the graphical process map. But as in the agile space, the model-driven development space is critically different from BPM RSD in one respect: it is meant to either generate source code or provide a framework within which source code is written. BPM RSD models are run-time models as well as design-time models and are designed to reduce the need to write custom software.

[2]Cantara, p.7.

So how are BPM RSD tools different? For agile or model-driven development, source code must be recompiled and redeployed when changes are made. Although this might be done automatically, it does not produce a clean transition in production environments; i.e., it is not seemless to an operating business process. BPM RSD products, however, store the semantics of process definitions in a repository – often a relational database – and execution engines dynamically drive production instances from this repository. Changes to the repository using the BPM RSD tools directly effect operational changes.

Another key distinction between frameworks and BPM RSD tools is that frameworks require skilled software developers to "wire up" the framework in order to achieve the benefits. Frameworks are analogous to e.POWER's API's plus our solution paradigm, but in e.POWER our framework has been pre-wired so that many of the technical requirements have already been resolved, allowing less skilled staff, or in some cases such as our process designer, non-technical staff, to contribute to solution development. Frameworks also require that individual wires be "soldered" into the solution. BPM RSD tools eliminate the need to do so and eliminate the possibility of *neglecting* to do so, insuring that critical functionality such as auditability, searchability, etc., mentioned in the Object Types section of this paper, is included automatically.

In a very real sense, BPM RSD tools are pre-wired, or pre-compiled frameworks.

IV. Out-of-the-Boxness

Similar approaches have arisen over the years in other business software categories. Typically packaged as products to offset the increased cost of producing these solution sets, these products consist of design tools that are largely configuration-driven and produce robust implementations. Such solution sets exist in the enterprise resource planning (ERP) space, customer resource management space (CRM), as well as the BPM space. Each of the design-time toolsets has unique characteristics. One of the key differentiators is how much functionality is delivered "out-of-the-box" and how much requires custom software development.

This "out-of-the-boxness" has significant benefits beyond the obvious advantage of creating solutions quickly. Successful BPM solutions must be customizable to each organization's unique requirements. Gathering those requirements through traditional documentation approaches can be cumbersome and slow and produces paper-based models to validate the requirements – an imperfect model at best.

BPM RSD tools provide working models of the solution in days rather than weeks or months. The significant user interfaces and workflow needed for requirements validation can be mocked up very quickly, providing a significant portion of the solution in a totally objective fashion – via working

ISSN 1947-5500

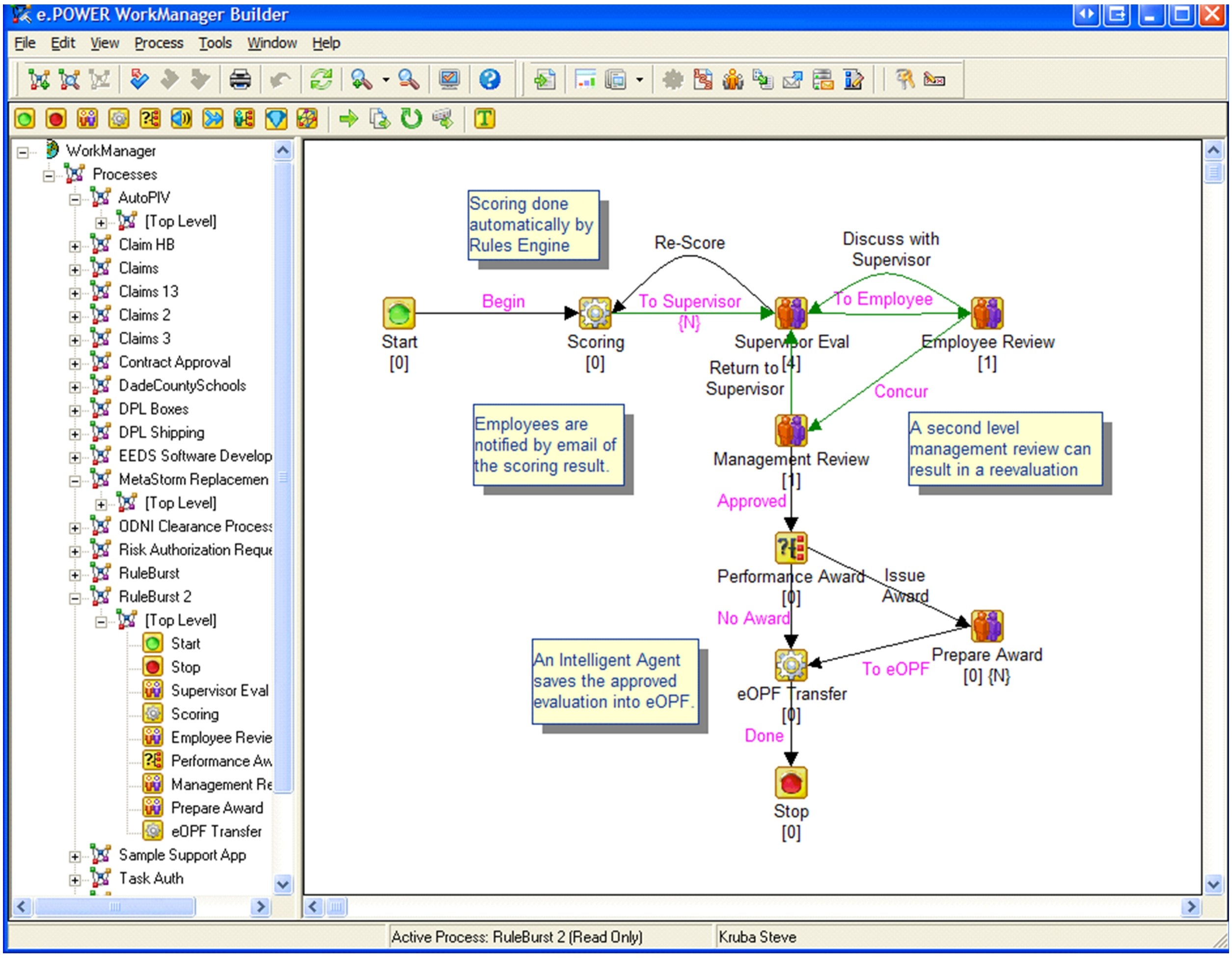


Fig. 2. Graphical Workflow Map

software. Select portions of the solution such as legacy systems integrations might be delayed to a later phase in order to minimize the impact on requirements gathering or to accelerate implementation in order to achieve operating efficiency gains earlier in the process. Iterating with prototypes help business people and IT staff objectify the end-product more quickly and accurately, greatly increasing the likelihood of a successful implementation.

The obvious advantage of BPM RSD development is also important: rapid implementations. Lengthy requirements efforts of many months suffer from the modern-day problem of a rapidly changing business context. How often have we seen a system that was well-designed and executed, but outdated by the time it was deployed? BPM RSD approaches reduce that risk.

It is important to note that these prototypes are not throwaways. To the extent they accurately reflect the underlying requirements, they become part of the final production solution. The key is that the tools used to develop proofs-of-concept, prototypes, and production systems are *the same tools.*

V. Object Types

Automation of business systems require creation of software modeling constructs that represent the key business objects in the problem domain. We refer to these constructs as *object types*. These object types map to the real-world business objects in the same way that classes relate to class instances in object-oriented programming languages. Object types could be thought of as index fields – or metadata on steroids.

Effective BPM RSD tools require a rich structure for creating process-enabled applications of any complexity. This generic structure, while necessary to support the user interfaces generated, is also very effective at helping analysts conceptualize the ultimate solution.

In tools such as the e.POWER Activator designer, object types define the characteristics of the real-world business components that make up the solution and object instances model the actual instances of those business objects. A by-product of this approach is that objects and object types inherit many useful properties from the e.POWER Activator infrastructure, features that might not be provided if the solution

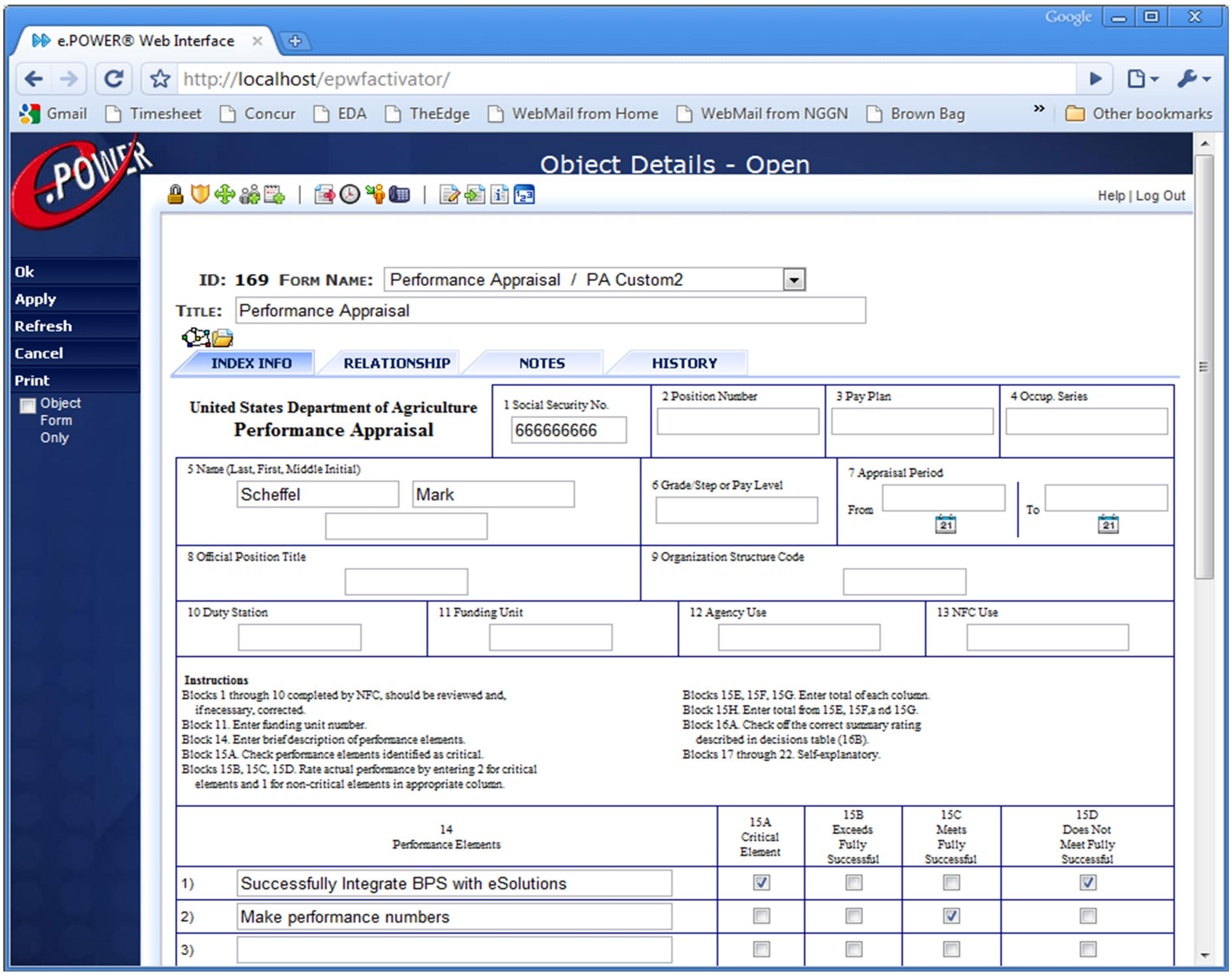


Fig. 3. User Interface

were created in a custom development effort. All e.POWER Activator objects are inherently editable and searchable and all activity against them is auditable. The rules applied at design time help to insure data integrity.

For an equal employment opportunity (EEO) application, object types might include complainant, class action lawsuit, and investigator. For a training solution, object type definitions would be needed for student, training class, and possibly training facility if the application was designed to model any of the facility behaviors. This direct mapping between IT constructs and business constructs greatly facilitate communication between IT staff and business staff and simplifies solution conceptualization.

VI. BPM RSD Benefits

Summarizing what we've discussed so far, BPM RSD tools provide the following key benefits over more traditional approaches.

❶ *Requirements gathering*. Providing a prototype solution early in the requirements gathering process helps end users understand the possibilities and helps to shape their expectations.

❷ *Requirements validation*. Working prototypes allow end users to understand exactly what they are getting. It is very difficult for end users to visualize how the software will affect how they work from paper documentation. Prototypes help to eliminate the perennial problem of "that's what I asked for but not what I need."

❸ *Analysis and design*. Prototypes also assist designers in visualizing what the ultimate solution can and should look like. BPM RSD capabilities allow them to draw from a toolbox of components that include "nice-to-have" features that might otherwise be omitted from the solution.

❹ *Documentation*. All solutions, whether using BPM RSD or more traditional approaches, require many forms of documentation: documentation for project approval, documentation for design reviews, documentation for the quality assurance process, etc. Having working prototypes early in the process makes all forms of documentation significantly easier to produce and much more effective. The clarity provided by actual screen-shots, process maps, and relational designs (necessarily generated automatically by BPM RSD tools) benefits all participants in the review

process, from end-users to the approving management staff.

❺ *Implementation*. Implementation is the obvious area where BPM RSD is valuable, allowing customers to achieve the benefits more quickly and less expensively than through traditional approaches.

❻ *Quality assurance*. The quality of a solution constructed through pre-built components is clearly higher since the out-of-the-box features have been refined through pre-existing, broad-based customer usage. New customers benefit from defects that were identified and resolved by other customers. Additionally, for each product release, the software is quality-checked through an independent process. This allows *project* quality teams to focus on the customizations – the area most likely to introduce software defects.

❼ *Maintenance*. The area of maintenance aligns with the notion of agility – being able to modify the production solution to adapt to changing conditions. The tools that facilitate rapid creation are typically the same tools used to update the solution as needs change over time.

❽ *Risk*. The BPM RSD approach reduces risk in virtually all phases of the development effort. The functioning prototypes reduce the risk of building a *good* solution that is the *wrong* solution. Analysis and design are improved through the objectivity of these same functioning prototypes, reducing the risk of an incorrect design. Quality is improved as noted above and therefore reduces the risk of poor quality. Implementation, operation and maintenance are likewise facilitated, reducing risk in their respective areas as well.

BPM RSD products affect all aspects of the system development life cycle and, as we shall describe in the next section, fundamentally change the way we approach solutions.

VII. A BPM RSD Methodology

As stated earlier, these new capabilities suggest a new approach to solution creation. [3] Rather than the traditional waterfall approach of requirements, design, development, and implementation, our approach is to use the following roadmap when engaging new customers. This approach is iterative: very similar to an agile software development approach, but the final result is achieved largely through model-manipulation rather than programming.

❶ Request existing documentation from the business users very early in the requirements gathering process.
  a) A Visio diagram or a description of the business process is the starting point for creating the process map.
  b) Copies of key forms provide templates for some of the user-interfaces as well as the data fields needed for the application.

❷ Prototype the solution using the BPM RSD tools.
  a) Use the graphical process designer to draw the business process which is more than visual: it encapsulates the business rules that drive the process. This graphical representation is critical to making sure IT and the business users agree on the process details.
  b) Use a declarative application builder for application creation.
  c) Use a security manager for defining security profiles, often integrated with an existing LDAP repository.

❸ Present the solution to the business users to refine the requirements and the solution.
  a) Iterate on changes to the process diagram in interactive design sessions.
  b) Modify the application in interactive design sessions.

❹ Put the solution into production, often in phases to accelerate the initial benefits.
  a) Create documentation from the prototype to support the organization's vetting process.
  b) Get something into production quickly to get immediate benefits.
  c) Defer complex integrations until later phases if possible.

This relatively simple formula is significantly more effective than alternative methods. Business users are able to react to high-fidelity prototypes rather than paper representations. BPM RSD tools make it feasible to rapidly evolve the solutions – interactively in design sessions with the end-users.

VIII. Conclusions

BPM rapid solutions development tools make it possible to construct business process solutions much more quickly and effectively than in the past. End-users are able to interact with designers in an expressive environment that allows them to model the solution with the actual tools used to create the solution. Rapid prototyping is a key to this approach and allows developers to build the solution that the business users need to satisfy their *actual* requirements – rather than the ones they "asked for."

This represents a major step forward in producing effective solutions. This approach results in lower risk of producing an ineffective solution and reduces defect rates by minimizing the amount of custom coding required to produce the solution. The end result is a much higher probability of successful projects. The BPMS software market serviced by model-driven BPM RSD tools may be unique in the IT software industry.

References

[1] D. Plummer and J. Hill, "Three Types of Model-Driven Composition: What's Lost in Translation?," Gartner, August 4, 2008
[2] M. Cantara, "Using the Four Corners Framework for BPM and BPM Usage Scenarios to Select BPM Consulting Vendors," Gartner, September 4, 2009
[3] D. Plummer and J. Hill, "Composition and BPM Will Change the Game for Business System Design," Gartner, December 21, 2009

Steve Kruba *is Chief Technologist for e.POWER product development and a Northrop Grumman Technical Fellow.*

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

A Hybrid Network Interface Card-Based Intrusion Detection System

Samir Elmougy,
Faculty of Computers and Information Sciences,
Mansoura University,
Mansoura 35516, Egypt,
mougy@mans.edu.eg

Mohammed Mohsen,
Faculty of Computers and Information Sciences,
Mansoura University,
Mansoura 35516, Egypt,
mohsen_cs@hotmail.com

***Abstract*—In recent years, the networks have played a vital factor in modern society. To prevent data tampering as well as eavesdropping, it's important to ensure that connections are always private and secure. Intrusion Detection Systems (IDSs) are gaining more importance to the applied technologies and become an integral part of the security infrastructure of organizations.**

In this paper, a new hybrid intrusion detection system called HSIDS combines both of heuristic and signature intrusion detection approaches is proposed and implemented based on reading bytes from the Network Interface Cards (NICs). Embedding the capturing module in the protocols stack is another capturing method used in HSIDS. HSIDS's structured is layered which allows to detect bugs fast and easily. Also, its functionality is not depending on any external applications, so it is easy to upgrade its protocols parsing classes. The experimentation results show that the proposed system is an efficient IDS.

Keywords-Computer security, hybrid intrusion detection system, network interface cards (NIC), heuristic intrusion detection, signature intrusion detection.

I. INTRODUCTION

Today, organizations rely on flexible and efficient security approaches and tools to guarantee that their information being exchanged is secured and privacy. Many approaches have been achieved to assure system privacy and security such as user authentication, authorization, encryption, firewalls, antivirus, and intrusion detection Systems (IDSs). Computer security is that field concerning with using technology, policies, and education to assure many factors such as the confidentiality, integrity, and availability of information system resources. This includes hardware, software, firmware, information/data and telecommunications [1, 2]. To secure data, three main activities should be pursued: prevention, detection, and recovery [3]. To be able to get a secure system, it is important to identity threats, extract characteristics from the threats, and encode the characteristics into software to detect those threats [4]. Intrusion is simply an attack attempting to access machine to get and/or manipulate information or to force it to be unreliable or unusable [5]. Intrusion can be unauthorized use, misuse, or abuse of computer systems by authorized user.

Firewalls are placed in between two or more computer networks to stop committed attacks into or out of these networks. Packet filtering firewall usually works by scanning a packet for both of the layer three and the layer four protocols information. A packet filtering firewall works by applying some filtering rules called policies. Provide information regarding whether the event is occurred or not cannot be obtained [2, 6, 7]. Firewalls are not totally enough to ensure the network security. Hence, intrusion detection systems (IDSs) are needed to identify malicious activity and suspicious in computer systems [8].

Intrusion detection systems depend on monitoring the computer systems or the networks to gather information, analyze this information, and recognize the system behavior to take a suitable action to prevent any completion of this attack and to ensure that the system is safe. IDSs are working by scanning packets at layer three and at layer four. IDSs can scan the different levels protocols of application and can also recognize the traffic type such as DNS, http and DNS [6]. IDS is alarming when there is a specific packet founded to match the parameters (the port number, the transport protocols (TCP/UDP), the IP address, the application protocols and the content) that are predefined by the IDS rules.

Two main methodologies namely anomaly detection and signature (misuse) detection are used in IDSs. Signature detection approach is effective for detecting those types of attacks without many false alarms. In the anomaly detection approach, the used heuristic function extends the power of the IDS dramatically since the admin will usually adjust it according to the very details of the network activities and nature. In other words, heuristic-based IDSs can cover all internal and external aspects of the network but signature-based IDS can cover only external aspects (attacks with signature). Heuristic based IDSs are limited only for attacks to exhibit abnormal behavioral patterns.

The main problems of using standard signature-based or anomaly-based IDSs is that their detection methods depend on detection instructions at the host processor level. Also, when an abnormal activity is detected using any of those

approaches, the anomalous packets will not be prevented from causing some bad effects such as trying to slow down or stop the system and the central processing unit. These problems cause the need to use Network Interface Cards (NICs) in the network intrusion detection applications [9, 10]. NICs are used to transfer data between different components of the system and the network. NIC first examines the transmitted packet headers and simply takes the decision of not forwarding any founded suspicious packets. Hybrid IDS is combined of two or more of IDSs architectures to overcome the drawback and weaknesses of using each one of these IDSs alone.

In this paper, a new intrusion detection system, we call it HSIDS, is proposed and implemented. HSIDS packet capturing depends upon reading bytes from the NIC by identifying the NIC system name in order to initialize handling for communicating with it. HSIDS combines both of heuristic detection and signature based detection approaches to overcome the drawback of using both alone.

This paper is organized as follows. In Section II, an idea about what is IDS, its types, methods, what it can do, and what it cannot do and discussing some related work are introduced. Our proposed system, HSIDS, is introduced in Section III. A discussion for how the package is captured using HSIDS is explained in Section IV. Section V covers HSIDS configurations and using. The conclusions and some future work are discussed in Section VI.

II. Background and Related Work

IDS system collects information from the networks and tries to detect attacks. It basically captures the flowing network stream of data and starts attempting to know if it threatens the network. IDSs types vary due to their methods of operations. Some common types of IDSs are:

1. Network IDS, NIDS: IDS that detects intrusions in a network
2. Distributed IDS, DIDS: IDS distributed on more than one host and may have a centralized log, analysis processing unit or an intrusion reporting unit (i.e. monitor).
3. Host IDS, HIDS: IDS that detects intrusions on a host (single workstation).

The place in a network to place IDS is greatly depending upon many factors as:

1. The purpose of the IDS: If the IDS is supposed to protect a whole network, then it should be seeing the whole network traffic. If it’s supposed to protect a node, then all that should be done is placing the IDS on that node. The main idea is just to see all the traffic needed. Adjusting the NIC filter is very important which it will be discussed later in "Capture a packet?” section.
2. Token of the Network: IDS is supposed to see all the traffic which it is supposed to check for intrusion signs. Assume that there is a ring token network then where should IDS is deployed? Deploying IDS in a ring token network is very expensive as the IDS will have to be able to see the traffic passing between every two nodes. So, usually the network structure is changed to permit efficient integration of IDS into the network.
3. The place of firewall: Assume that there is a network sees the internet through a firewall that acts as a bottleneck to the network connection. An ideal place to deploy the IDS is where the data stream is supposed to be filtered. In other meaning, IDS should be placed according to the diagram given in Fig. 1.

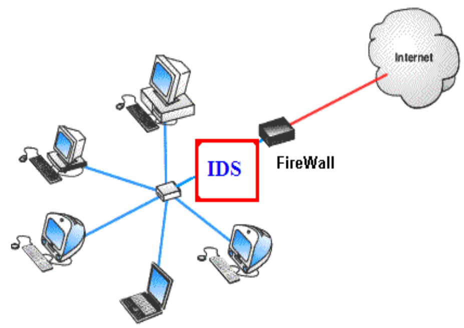


Figure 1. Positions of IDS and Firewall

4. Mistakes usually occurred when deploying IDS: The following are some mistakes usually occurred when deploying IDS systems [12]:
   - Deploying the network IDS without sufficient infrastructure planning.
   - When the IDS is deployed appropriately, but nobody is looking at the alerts it generates
   - Network IDS is deployed, "sees" all the traffic and there is a moderately intelligent somebody reviewing the alert stream.
   - All the previous pitfalls are avoided and the NIDS is humming along nicely. However, the staff monitoring the IDS starts to get flooded with alerts.
   - Not accepting the inherent limitations of network IDS technology. While anomaly-based IDS systems might potentially detect an unknown attack, most signatures based IDS will miss a new exploit if there is no rule written for it.

IDS alerts have a ratio of falseness and needs adjustments. The alert reporting method is significant, whether it will send a mail, pop up a message, and start a sound declaring an attack or even send an SMS to the network administrator. Many IDSs can only analyze the attacks but others try to stop the attack at the time of the

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

intrusion. Network traffic data, system status files, system level test data, are the main types of data used by IDSs [13].

Two main different methodologies in designing intrusion detection systems are signature-based and Heuristic-based. Heuristic-based (synonymous with anomaly-based) IDSs approach deal with the uncovering the behaviors of abnormal patterns given a model of user's normal behavior. So, any event causes violating the model is a suspicious. This usually implies the use of extensive attack free training sets in order to characterize normal behavior. The alerting phase comes when a pre defined level of deviation occurs. If some protocols start taking over the bandwidth, the bandwidth availability is running low, so many login failure on a specific machine. When a huge deviation occurs from the usually snap shot of the network, alert is issued. Anomaly detection is very powerful for detecting DoS attacks, network scanning and sniffing, but it could be easily fooled. A simple attack needing no more than launching an exploit won't be an enough deviation from the original state of the network. However, it has the drawback of producing high false alarms if a reasonable suspicion level is not maintained. Statistical approaches such as PHAD [14] IDS, Finite mixture model [15], clustering and data mining [16], artificial neural networks [17], Expert Systems such as MIDAS, IDES and NIDES, genetic algorithms such as the IDS given in Crossbie [4], machine learning and immune systems techniques are the main categorizations of anomaly detection systems.

Signature detection which is called also misuse or detection by appearance systems rely on the use of specific known patterns of unauthorized behavior and/or contents (parts of the attack signature). This technique is fast and very accurate when it comes to detect a specific attack because it checks the protocol layers for known signatures. Encoding can fool signature based attacks but this usually applies only to web applications attack like cross site scripting and SQL injections. However, it has the drawback of possibility failure in detecting novel attacks whose signatures are unknown or in the case of environment changes. Snort [18] is an IDS running over IP-networks and depending on the signature-based intrusion detection system approach [19, 20].

Because a home-network-node cannot send a packet to itself from out of the network and a connection cannot be initiated from the port Zero, heuristic intrusion detection methods mainly depend upon the admen's past experience and intelligence. This type extends the power of the IDS dramatically since the admin will usually adjust it according to details of the network activities and its nature. One of the disadvantages is that bad rules will raise lots of false alerts which may lead to ignore alerts while an alert be a positive one. So access care should be taken when coding heuristic rules.

NIC is used to move data through the different system components and the network. It first examines the transmitted packet headers and simply takes the decision of not forwarding any founded suspicious packets. IDSs based on NICs can result in better performance of the overall network security system because NICs can provide IDS by [9, 11]:

- Better coverage: a one-to-one mapping between NICs and hosts.
- Scalability: natural distribution of computation.
- Less aggregation: detect more specific intrusions.
- Detecting intrusion internal to a LAN
- Potentially detecting more complex exploits by cooperating NICs.
- Improving performance by independency from host adds to reliability.

The overall architecture for NIC-based security is shown in Fig. 2 [9].

A P(srcIP | destIP) framework of is an example of anomaly IDS implemented based on the firewall and host NICs [21, 9]. A distributed version of P(srcIP | destIP) known as P(src IP | destIP,destPort) is implemented on the host NIC [9]. Embedding the firewall-like security at the NIC level is given in [8].

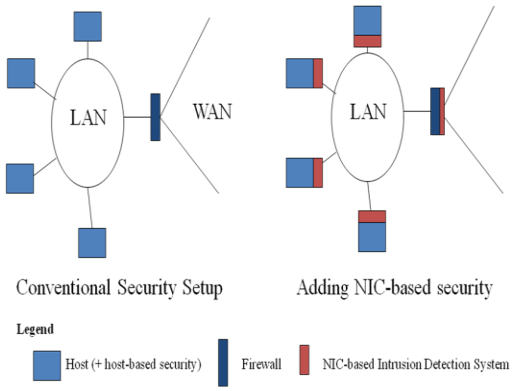


Figure 2. The architecture for NIC-based security

Weinsberg et al. [11] implemented a SCIRON (Secure-Communication IntegRated over NIC) firewall based on a NIC. Schuff et al. [22] presented and implemented a NIC-based IDS based on the processing of the available resources in future multi-core RISC processors combined with specialized content inspection hardware. Using Myrinet cluster to design and implement NIC-based QoS is presented in [23]. In 2001 [24], Markham et al. and Payne proposes and implemented a distributed firewall on a NIC. Sekar et al. designed

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

and implemented a hybrid IDS of anomaly detection approach with human-designed state machine [25].

Tombini et al. [26] combined signature and anomaly detection techniques to design and implement a hybrid IDS. Aydın et al. proposed a hybrid IDS combined of anomaly-based IDSs and network traffic anomaly detection (NETAD) based on the misuse-based IDS Snort [19].

III. THE PROPOSED HYBRID INTRUSION DETECTION

The proposed IDS system, HSIDS, is modified using a Pacanal package, a winpcap C# mimic. The will known winpcap library [4] had been translated into C#. In this package, an ethereal-like application depending on winpcap technology implemented using C# is implemented with supporting APIs. Pacanal was just a packet capturer and needed an enormous amount of effort to develop. For Pacanal package, there is no need to send any packet although its designer implemented the Winsock service initialization and an API function is used to write byte arrays into the NIC directly which could be used to craft packets.

Pacanal package's power is extended but meanwhile all unneeded functions and protocol parsing classes are removed. Pacanal's configuration panel has many options regarding being a packet capturer configuration panel. But HSIDS's configuration panel is hanged with about 85%. HSIDS is capable of working on almost all windows computers including the following versions (WIN2000, WINXP, WINVISTA, WINNT, WIN95, WIN98, and WINME).

In our proposed system, HSIDS, the packet capturing depends upon reading bytes from the NIC. This method is depending on the identifying the NIC system in order to initialize a handle for communicating with it. In order to capture a packet, the current NIC is identified and its parameters are specified. HSIDS's structured is layered which allows to detect bugs fast and easily. Also a great ease in upgrading HSIDS is achieved. Moreover, HSIDS's protocols parsing classes could be increased and integrated into the project very easily. The following algorithm shows how HSIDS is working.

1. Reading packet from NIC.
2. Parsing packet initially using the frame parser and Ethernet protocol parser.
3. Ethernet protocol parser parses the standard fields for a typical Ethernet header and also identifies the upper protocol whether it is TCP, UDP, ARP, etc.
4. According to the detected protocol, the appropriate packet parsing class parsing function is called and the rest of the packet is passed to that class function.
5. IDS for each protocol is present in the shape of classes named as follows udpIDS.cs, tcpIDS.cs … etc.
6. In each protocol parsing class, a module from the relevant IDS class is called to detect possible intrusion signs.

Any protocol parsing class could be easily added and integrated in the appropriate protocol layer (e.g. after transport a protocol for example).

As mentioned early, HSIDS depends upon the capturing infrastructure of Pacanal which depends mainly in itself for capturing packets and raising the obtained byte to the upper layers of HSIDS for parsing and intrusion detection. Although, winpcap libraries when setup it extends HSIDS's reliability by assuring existence of the npf.sys driver as an example. Some HSIDS bugs are avoided when installing WinPcap.

Signature-detection IDSs used to detect known attacks but anomaly detection IDSs can detect new attacks methods of heuristic. HSIDS is implemented using both of signature-based and anomaly-based (by using a heuristic function to extend the power of the IDS) intrusion detection approaches. Capturing a packet is a little complicated process and many steps should be made before starting to capture a packet. Similar to winpcap, Pacanal's descent which is HSIDS uses the easiest way of packet capturing. It simply reads packets from the NIC. So, it's counted as a protocol to read packets from the NIC.

Another method of capturing is to embed the capturing module in the protocols stack, so that the packet should pass by the capturing module and this capturing should pass it to the upper protocol layer depending on where the capturing module is added in the protocol stack. This method can choose to pass or not to pass the packet received from the lower protocol layer. Also, this method show how most of the firewalls can be worked and also how some IDSs, that increases their features by such an option, discard specific type of packets.

A. *Identifying the Platform*

In order to capture a packet, the current NIC in use is identified first followed by specifying its parameters. A packet32h object is created and when created it:

1. Get the operating system info.
2. Get the list of up and working network adapters.
3. Initialize the winsock.dll.

To identify the current windows version, certain API functions are called and variables are passed by reference in order to send the variable and receive it again with its values. The API function is:

```
[DllImport("kernel32.dll")]
public extern static int GetVersionEx(ref
OSVERSIONINFOmOSInfo );
mOSInfo.dwMajorVersion &
mOSInfo.dwMinorVersion
```

The mOSInfo is a struct that has many variables in it.GetVersionEx API function previously know that it will receive a variable with that structure.

B. Opening the NIC

This API function opens the NIC using its previously obtained system name as a file for read and write access modes, and creates it depending on that it already exists. Also, it can write packets bytes in the NIC and eventually injecting crafted packets into the network. For example, the WinXP which is our OS lies under the Win2000 category so to find the list of current network interface cards, the list of keys is checked in the following registry path:

SYSTEM\\CurrentControlSet\\Control\\Class\\{4D36E972-E325-11CE-BFC1-08002bE10318}

After receiving the NIC's system known name, a check is required on the device system name to make sure it obeys the following format: "\Device\NPF_{TheDeviceSystemName}". The following step is to call the following function:

```
[DllImport("kernel32.dll")]
public extern static int CreateFile (
char [] lpFileName,
/* pointer to name of the file Device system    name*/
int dwDesiredAccess
/* access (read-write) mode Read and Write  */
int dwShareMode,          /* share mode 0 */
int lpSecurityAttributes,
/* pointer to security attributes 0        */
int dwCreationDistribution,
/* how to create 3 "Open existing“ */
int dwFlagsAndAttributes,
/* file attributes 0 */
int hTemplateFile); * handle to file with attributes to
copy 0
```

This function returns an integer which is the NIC's handle that will be used to deal with the NIC's I/O stream in the memory. Another API function of the kernel32.dll is:

```
[DllImport("kernel32.dll")] public extern static int
DeviceIoControl(
int hDevice, uint dwIoControlCode,
uint [] lpInBuffer, int nInBufferSize,
int lpOutBuffer, int nOutBufferSize,
ref int lpBytesReturned, int lpOverlapped
 )
```

This helps to set attributes to the device with the specified handle or issuing commands to the device. This function has eight different overloads to serve that issue.

C. Reading a Single Packet

The following function issues a command to the NIC to make one read operation.

```
[DllImport("kernel32.dll")]
public extern static int WaitForSingleObject
( int hHandle, uint dwMilliseconds );
```

This function actually reads the object (byte[] packet) obtained from the NIC.

```
[DllImport("kernel32.dll")]
private static extern bool ReadFile (
int hFile,                    // handle to file
byte [ ] lpBuffer,            // data buffer..output
int nNumberOfBytesToRead, //number of bytes to
read
ref int lpNumberOfBytesRead,  // number of bytes
read
ref OVERLAPPED lpOverlapped   // overlapped
buffer
);
```

D. Mess Cleaning

First, mess should be cleaned and free all system resources that were reserved by HSIDS using the following function to end the NIC commands session. For example:

```
[DllImport("kernel32.dll")]
public extern static int CloseHandle
( int hObject ); //The NIC's handle
```

IV. HSIDS Configuration and User Interface

Logs are saved in .mdb access db format in the ".\Logs" directory. A log file is named after the time and the time and date the HSIDS started capturing packets. An example of log file is shown in Fig. 3.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

sip	dip	Sport	dport	Sign	Msg	References	type
Any	any	Any	135	7416e877cce0fd7fcce0fd7f	DCOM Exploit (MS03-026)	www.microsoft.com/security/security_bulletins/ms03-026.asp	admin
Any	any	Any	135	ec29e877cce0fd7fcce0fd7f	DCOM Exploit (MS03-026)	www.microsoft.com/security/security_bulletins/ms03-026.asp	admin
Any	any	any	135	b524e877cce0fd7fcce0fd7f	DCOM Exploit (MS03-026)	www.microsoft.com/security/security_bulletins/ms03-026.asp	admin
Any	any	any	135	7a36e877cce0fd7fcce0fd7f	DCOM Exploit (MS03-026)	www.microsoft.com/security/security_bulletins/ms03-026.asp	admin
Any	any	any	135	9b2af977cce0fd7fcce0fd7f	DCOM Exploit (MS03-026)	www.microsoft.com/security/security_bulletins/ms03-026.asp	Admin
Any	any	any	135	e3afe977cce0fd7fcce0fd7f	DCOM Exploit (MS03-026)	www.microsoft.com/security/security_bulletins/ms03-026.asp	Admin
Any	any	any	135	ba26e677cce0fd7fcce0fd7f	DCOM Exploit (MS03-026)	www.microsoft.com/security/security_bulletins/ms03-026.asp	admin

Figure 3. An example of log file

A. *User Interface*

Fig. 4 shows a screen shot of the main user interface of HSIDS. The main parts that are appeared in this figure are the main menu and five main windows as follows. The menu items are divided into two options (Capture which is indicated by the number "1" and Options which is indicated by the number "2"). "Capture" option is used either to start the capturing process through using the option "Start", which is indicated by the number "3", or to stop capturing through using "stop" option which indicated by the number "4". The menu item "Options" which is indicated by the number "2" and is used either to change HSIDS configuration in addition to getting some help through "Configure HSIDS" option, which is indicated by the number "5", or to exit the system through "Exit" option which is indicated by the number "6".

The following are the five main windows that are appeared in Fig. 4.

- "Tree View" indicated by the number "7": It shows a tree structure for a shown packet holding a threat.
- A rich text box control indicated by the number "8": It shows the HEX dump for a shown packet holding a threat
- A rich text box control indicated by the number "9": It shows information about the threat, how to deal with, and what is usually provided.
- A rich text box control indicated by the number "10": It shows statistics about protocols, amount of bytes and time elapsed.
- A list box control indicated by the number "11": It shows a list containing a brief description about the protocol, threat, packet ID and time of arrival
- A label control indicated by the number "12": It shows HSIDS's slogan.

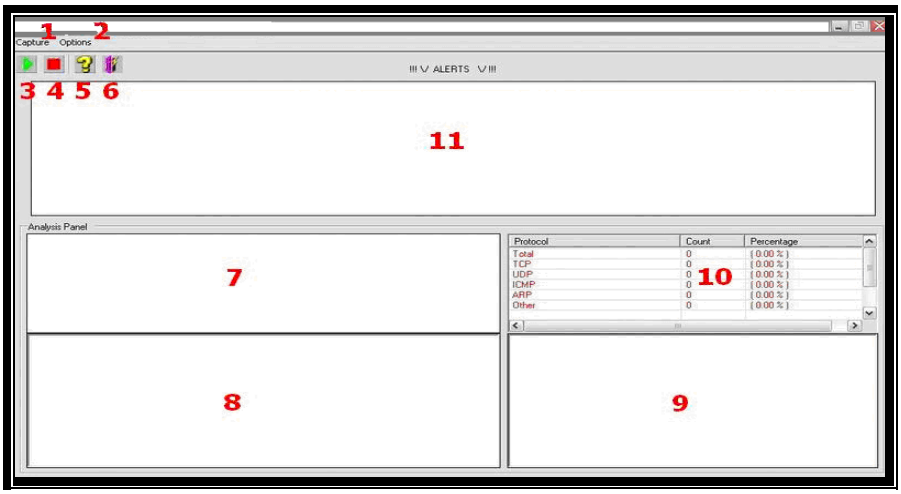


Figure 4. A screen shot of HSIDS system.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

B. Signature DB

HSIDS has a signature database (DB) for many know attacks. Signatures are stored in many databases with relation to the protocol itself for example TCP has a DB for all its types of attacks (tables) and each table has its own rules sets containing a signature for the attack. When an attack is launched, the attacking packets will have some fingerprint or a signature that declares its threat.

An example of HSIDS's signature rules is given in Fig. 5. In this figure, the column "sign" represents hex strings. IF found as a TCP payload coming from any IP going to any IP from any port to port 135 then this is the well known. This method is very accurate when it comes to detecting a specific attack because it checks the protocol layers for known signatures

C. Heuristic-Based Intrusion Detection

Heuristic intrusion detection depends mainly on how a strange behavior would be. An IP IDS heuristic module is given as:

```
private void heuristic()
{
Int32 int1=0;
if((astn.LocalIP()==astn.SIP())
||(astn.SIP()==astn.DIP()))
//Unlogical source and destination IPs {
    / /Logging a possible unsecure header
    cmd_.CommandText=
    "insert into unsecure (pid,protocol,sign) VALUES
('"+pid+"','IP','Un logical source and target IPs')";
    int1 = cmd_.ExecuteNonQuery();
    //Reporting strange activity.
    lstbox.Items.Add("[IP][Heuristic Scan][Un logical
source and destination IPs]]
#"+Convert.ToString(Convert.ToInt32(pid)-
2)+"at"+DateTime.Now.TimeOfDay.ToString());
   conn_.Close();astn.CloseConnection();
  } }
```

D. HSIDS Configuration Panel

HSIDS is capable of copying any packet that passes by the NIC of the host having HSIDS running on it. HSIDS obtains the packet in a byte [] format and can efficiently parse the array. As mentioned before, HSIDS opens a NIC with read and write access modes which means that HSIDS can craft. Pacanal's configuration panel had many options regarding being a packet capturer configuration panel. HSIDS is capable of working on almost all windows computers including the following versions (WIN2000, WINXP, WINVISTA, WINNT, WIN95, WIN98, and WINME). A screen shot is given in Fig. 6 to show the main HSIDS's configuration panel where:

- The NIC device name is indicated by the number 1 in the interface.
- An option to limit the number of data captured of each packet is indicated in the interface by the number 2.
- An option to limit the number of packets captured for intrusion detection is indicated in the interface by the number 3. An option to limit the number of kilobytes captured for intrusion detection is indicated in the interface by the number 4.
- An option to limit the time elapsed during intrusion detection is indicated in the interface by the number 5.
- An option to specify the buffer size of the NIC is indicated in the interface by the number 6.
- An option to specify the buffer size of the intrusion detection is indicated in the interface by the number 7.
- An option to specify how much data should the HSIDS copy from the NIC's buffer for intrusion detection (the minimum amount of data needed to copy in each read process from the NIC's buffer) is indicated in the represented by number 8.

sip	dip	sport	dport	sign	msg
any	any	any	135	7416e877cce0fd7fcce0fd7f	DCOM Exploit (MS03-026)
any	any	any	135	ec29e877cce0fd7fcce0fd7f	DCOM Exploit (MS03-026)
any	any	any	135	b524e877cce0fd7fcce0fd7f	DCOM Exploit (MS03-026)
any	any	any	135	7a36e877cce0fd7fcce0fd7f	DCOM Exploit (MS03-026)
any	any	any	135	9b2af977cce0fd7fcce0fd7f	DCOM Exploit (MS03-026)
any	any	any	135	e3afe977cce0fd7fcce0fd7f	DCOM Exploit (MS03-026)
any	any	any	135	ba26e677cce0fd7fcce0fd7f	DCOM Exploit (MS03-026)

Figure 5. HSIDS signature

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

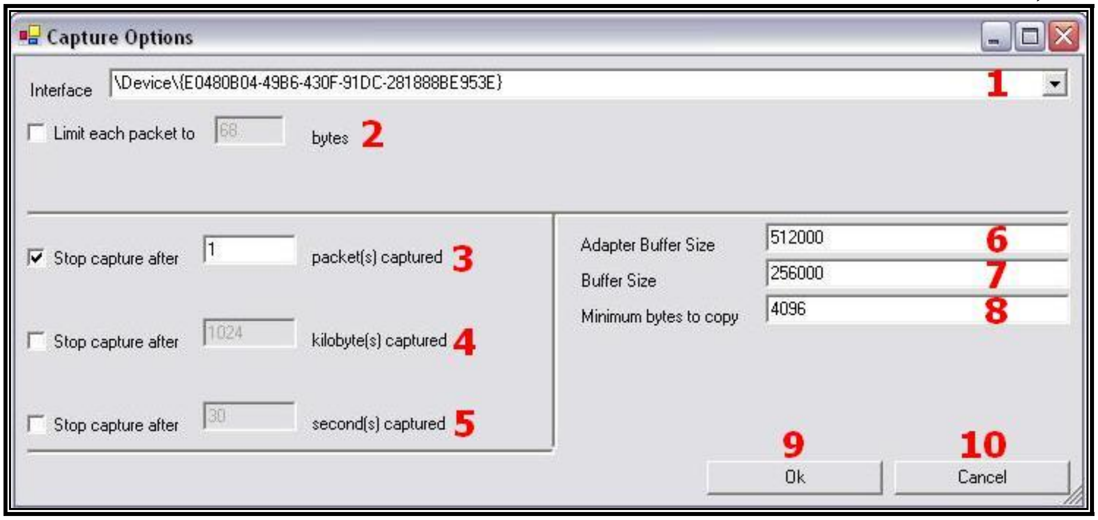


Figure 6. A screen shot HSIDS configuration panel

- A button to save and apply the options is represented in the interface by the number 9.
- A button to cancel the configuration screen and return back to the main interface is represented in the interface by the number 10.

E. Maintaining Order when Discovering an Attack

If a spoof attack is launched a primary step to deal with the attack to launch fake packets that acts as a spoofer, it returned everything that been used even spoof the attackers IP and cutting it of the network. Although bypassing a spoofed attack is very easily even manually, it is the least thing we can do as a favor to the attacker.

F. HSIDS's Mutilation

In HSIDS, it is also implemented how to switch to the stack based capturing method to provide more options mainly preventing some packets from passing through, mutating HSIDS, and turning it into a hybrid IDS/IPS solution.

V. PRELIMINARY EXPERIMENTS

IDS validating is important to measure its performance. For preliminary experimental study, two victim machines running on Windows XP operating systems are used for the experimentation. The traffic generators of other hosts machines and different users who are using different applications and internet are simulated.

A set of validating data is gathered from the two victim machines and from the network. First, we trained anomaly detection systems to one of the following attacks categorizations: (Probing “PRB”, User to Root “U2R”, Denial of Service “DOS”, Remote to Local “R2L”) as shown in Table I. The following step is to provide the test data containing 92 unlabeled instances of attacks without predefining 22 of these attacks in the training data stage.

TABLE I. THE PERCENTAGES OF THE DIFFERENT CATEGORIZATION OF ATTACKS OF THE TRAIN AND TEST DATA

Attack Categorization	Train Ratio	Test Ratio
Normal	42.0%	22.36%
PRB	17.0%	3.43%
U2R	7.0%	1.19 %
DOS	30.0%	64.21%
R2L	4.0%	8.82%

To apply the validating measures on the experimental results, Table II lists the parameters required for these measures.

TABLE II. THE USED PARAMETERS IN THE SYSTEM VALIDATING PROCESS

Parameter	Parameter symbol	Definition
True Positive Rate	TP	Attack occurs and in the same time alarm raised
True Negative Rate	TN	No attack occur and in the same time no alarm
False Positive Rate	FP	No attack occur and no alarm raised in the same time
False Negative Rate	FN	Attack occurs and no alarm raised in the same time

Table III shows the final results using the following measurements to validate the performance of HSIDS [27]:

- Precision measure: It represents the occurring of an attack and in the same time this attack is correctly detected. It is computed as:

$$\text{Precision} = TP / (TP+FP).$$

- Recall measure: It represents the occurring of an attack and in the same time detecting attacks from the really attacks. It is computed as:

$$\text{Recall} = TP / (TP + FN)$$

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

- Detection Rate: It represents the ratio between the total attack number and the total detecting number of attacks.
- The false alarm measure: It represents the occurring of attack and in the same time the system could not correctly detect it or the attack happens. It is computed as:

The false alarm = $(FP + FN) / (TP + FP + FN + TN)$

TABLE III. THE FINAL RESULTS OF THE SYSTEM VALIDATING

Categorization	Detection Rate	False Alarm	Precision	Recall
Normal	95.19%	4.81%	88.24	98.21
PRB	96.78	3.22%	83.43	88.81
U2R	84.65%	16.35%	78.94	74.3
DOS	97.62%	2.38%	98.12	98.54
R2L	61.02%	38.98%	83.22	10.41

From the results, it is shown that the HSIDS is suitable for detecting errors that are predefined and not predefined in the database. Also, it can achieve a very good overall accuracy in detecting attacks.

VI. CONCLUSION AND FUTURE WORK

It's very obvious that IDSs are gaining more importance by the day due to the used applied technologies applied through it regarding to the respond to attacks, and the capability of identifying the origin of these attacks. High data flow rate is a ruthless enemy and may greatly affect the performance of IDS, especially large packets.

In this paper, a new hybrid IDS called HSIDS in which its capturing capability depends upon reading bytes from the NIC is proposed and implemented. Its capturing method depends on embedding the capturing module in the protocols stack so that the packet can be passed by the capturing module to the upper protocol layer depending on where the capturing module is added in the protocol stack. In other meaning, HSIDS combines heuristic and signature based detection approaches. HSIDS's structured is layered which improves its capabilities in detecting bugs fast and easily. It is easy to upgrade HSIDS's protocols parsing classes and integrate it into most of other projects in very easily matter because it does not depend on any external applications.

HSIDS is tested itself by giving infrastructure to craft fake packets then launching fake packets towards HSIDS where HSIDS succeeds in detecting the attack embedded in the packet. HSIDS is tested through an experimental study where the results show that it is suitable for detecting errors that are predefined and not predefined in the database with achievement a very good overall accuracy in detecting attacks.

As a future work, we plane to investigate the performance of IDS in details using a suitable database of attacks and compare its performance efficiency with other IDSs under different conditions.

REFERENCES

[1] Bishop, M., *Computer Security: Art and Science*, Addison-Wesley, Boston, MA, 2003.
[2] Seymour Bosworth, M.E. Kabay, *Computer Security Handbook*, 4th ed., John Wiley & Sons, 2002.
[3] Marcus A. Maloof, *Machine Learning and Data Mining for Computer Security: Methods and Applications*, Springer-Verlag London Limited, 2006.
[4] Philip K. Chan, Richard P. Lippmann, "Machine Learning for Computer Security," *Journal of Machine Learning Research*, vol. 7, pp. 2669-2672, 2006.
[5] Sathish Alampalayam P. Kumar, Anup Kumar, and S. Srinivasan, "Statistical Based Intrusion Detection Framework using Six Sigma Technique," *IJCSNS International Journal of Computer Science and Network Security*, vol.7, no.10, October 2007.
[6] (2003) Joe Bowling, "The Future of IDS". [Online]. Available: http://www.infosecwriters.com/texts.php?op=display&id=115
[7] http://www.winpcap.org/docs/docs31/html/group__NPF.html
[8] Bace R.G., "Intrusion Detection," Indianapolis, USA, Macmillan Technical Publishing, 2000.
[9] M.Otey, R. Noronha, G.Li, S. Parthasarathy, and D. Panda, "NIC-based Intrusion Detection: A feasibility study," Proceedings of the IEEE ICDM Workshop on Data Mining for Cyber Threat Analysis, December 2002.
[10] M. Otey, S. Parthasarathy, A. Ghoting, G. Li, S. Narravula, and D. Panda, "Towards NIC based intrusion detection," in Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 723–728. ACM, ACMPress, NY, USA, Aug. 2003
[11] Yossi Amir, Gilad Gat, Elan Pavlov, Yaron Weinsberg, Sharon Wulff, "Putting it on the NIC: A Case Study on application offloading to a Network Interface Card," Consumer Communications and Networking Conference CCNC 2006.
[12] O1-Anton Chuvakin, Five IDS MisHSIDSes People Mak. [Online]. Available: http://www.computerworld.com/securitytopics/security/story/0,10801,78670,00.html?SKC=security-78670
[13] Bace R.G., "An introduction to intrusion detection and assessment for system and network security management," ICSA Intrusion Detection Systems Consortium Technical Report, 1999.
[14] Matthew V. Mahoney and Philip K. Chan, "PHAD: Packet header anomaly detection for identifying hostile network traffic," Technical Report, Florida Tech., 2001.
[15] K. Yamanishi, J. Takeuchi, G. Williams, and P. Milne, "On-line unsupervised oultlier detection using finite mixtures with discounting learning algorithms," In KDD, pages 320–324, Boston, MA, 2000.
[16] Eleazar Eskin, Andrew Arnold, Michael Prerau, Leonid Portnoy, and Sal Stolfo., "A geometric framework for unsupervised anomaly detection: Detecting intrusions in unlabeled data," Data Mining for Security Applications, 2002.
[17] Jake Ryan, Meng-Jang Lin, and Risto Miikkulainen, "Intrusion detection with neural networks," In Proceedings of AAAI-97 Workshop on AI Approaches to Fraud Detection and Risk Management, pages 72–77, AAAI Press, 1997.
[18] (2010) homepage of Snort. [Online]. Available" http://www.snort.org/

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[19] M. Ali Aydın, A. Halim Zaim, K. Gokhan Ceylan, "A hybrid intrusion detection system design for computer network security," Computers and Electrical Engineering 35, 517–526, 2009.
[20] Roesch M., "Snort – lightweight intrusion detection for networks," In Proceedings of the 13th LISA Conference of USENIX Association, 1999.
[21] M. Mahoney and P. Chan, "Learning nonstationary models of normal network traffic for detecting novel attacks," In SIGKDD, 2002.
[22] D. Schuff, V. Pai, P. Willmann and S. Rixner, "Parallel Programmable Ethernet Controllers: Performance and Security," IEEE Network, 2007.
[23] A. Gulati D. K. Panda P. Sadayappan and P.Wyckoff, "NIC-based rate control for proportional bandwidth allocation in myrinet clusters," In Int'l Conference on Parallel Processing, 2001.
[24] Markham, T. and Payne, C., "Security at the network edge: a distributed firewall architecture," In DARPA Information Survivability Conference & Exposition II, 2001.
[25] Sekar, R., Gupta, A., Frullo, J., Shanbhag, T., Tiwari, A., Yang, H., and Zhou, S., "Specification-based anomaly detection: a new approach for detecting network intrusions," In Proceedings of the 9th ACM conference on Computer and communications security, ACM Press, pp. 265–274, 2002.
[26] Tombini, E., Debar, H., M´E, L., and Ducass´ E, M., "A serial combination of anomaly and misuse IDSes applied to HTTP traffic," In 20th Annual Computer Security Applications Conference, 2004.
[27] G. Helmer, J.S.K. Wong, V. Honavar, and L. Miller, "Automated discovery of concise predictive rules for intrusion detection," Journal of Systems and Software, Vol. 60, Issue 3, pp. 165–175, 2002.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Scheduling of Workflows in Grid Computing with Probabilistic Tabu Search

R. Joshua Samuel Raj
CSE, VV College of Engineering
Tirunelveli, India
joshuasamuelraj@gmail.com

Dr. V. Vasudevan
Prof. & Head/IT, Kalasalingam University
Srivilliputur, India
drvvmca@yahoo.com

Abstract:

In Grid Environment the number of resources and tasks to be scheduled is usually variable and dynamic in nature. This characteristic emphasizes the scheduling approach as a complex optimization problem. Scheduling is a key issue which must be solved in grid computing study and a better scheduling scheme can greatly improve the efficiency.The objective of this paper is to explore the Probabilistic Tabu Search to promote compute intensive grid applications to maximize the Job Completion Ratio and minimize lateness in job completion based on the comprehensive understanding of the challenges and the state of the art of current research. Experimental results demonstrate the effectiveness and robustness of the proposed algorithm. Further the comparative evaluation with other scheduling algorithms such as First Come First Serve (FCFS), Last Come First Serve (LCFS), Earliest Deadline First (EDF) and Tabu Search are plotted.

Key words: grid computing, workflow, Tabu Search, scheduling problem, Probabilistic Tabu Search

INTRODUCTION

Grid Computing a pioneer technique in harnessing the geographically dislocated computer power has changed the perception on the utility and availability of the computer power, which has carved a new technology that openly ventures and amalgamates an infinite number of computing devices into any grid environment, augmenting to the computing capability and providing resolutions to the various tasks within the operational grid environment basically by enabling, sharing, selection and aggregation of geographically distributed autonomous resources dynamically at runtime, depending on their availability, capability, performance and cost, thereby shifting the focus to collaborative environments, federating services and exchanging transactions in a mutual manner to share resources and thereby achieve common goals to enhance productivity and speed up progress in much the same way that the Internet did in yesterdays economy, paving the way for numerous research efforts in grid scheduling mechanisms

Grid Computing is our greatest hope for delivering computing as utility to homes and offices. Many large scale applications such as scientific, engineering and business problems (Hai *et al.,* 2005; Cannataro *et al.,* 2002) are solved effectively using the logical amalgamation of geographically dispersed Grid resources (Bernan *et al.,* 2002). Grid computing, analogous to the pervasive electrical power grid, enables resource sharing and cooperative work among distributed computational sites. In grid environment, applications are often described as workflows. A workflow is composed of atomic tasks that are processed in specific order to fulfill a complicated goal. Generally, grid workflows require huge intensive computing and process larger data, compared with traditional workflows. Therefore, the performance of grid workflows becomes a critical issue of the workflow management systems. One of the most challenging problems is to map each task to a corresponding service instance to achieve the customers' quality of service (QoS) requirements as well as to accomplish high performance of the workflow. This problem is found to be NP-complete. During the course of grid scheduling there are many challenges that require the simultaneous optimization of several incommensurable and competing objectives.

- Unpredictable challenges in Grid resources
- Inevitability to multiple resource types for completing a job
- Necessitate for a parallel or concurrent execution of tasks in any workflows.

Under the OSGA, the workflow scheduler has to balance several QoS requirements, including makespan and cost. Consequently, many traditional workflow scheduling algorithms, such as Opportunistic Load Balancing, Minimum Completion Time, Min-min, Max-min and Duplex, are not

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

suitable since they only tackle the makespan requirement.

In recent years, a number of researches have been focused on scheduling problem involving more than one QoS requirements. The traditional System namely advanced reservations for scheduling the workflows undergoes problems such as overloading and power failure. The overloading and the scheduler failure problem are overridden by a two level scheduling scheme where the first level is used for frequent small jobs and second level for large jobs. The market oriented approach algorithm succeeded in distributed scheduling of workflows, but could not appease completion of more workflows within the deadline. The success ratio of the workflows allotted for mapping the Grid sites is 30% (Chien et al., 2005) when 30 workflows are scheduled at a time.

Workflows submitted to the Computational Grids by resource consumers have a proper budget proposal, client authentication and the requirements for its execution as shown in Fig 1. The willingness to complete any job is given by resource providers. Hence the Grid schedulers search for solutions in the state space aiming at achieving high performance, both in terms of solution quality and execution speed.

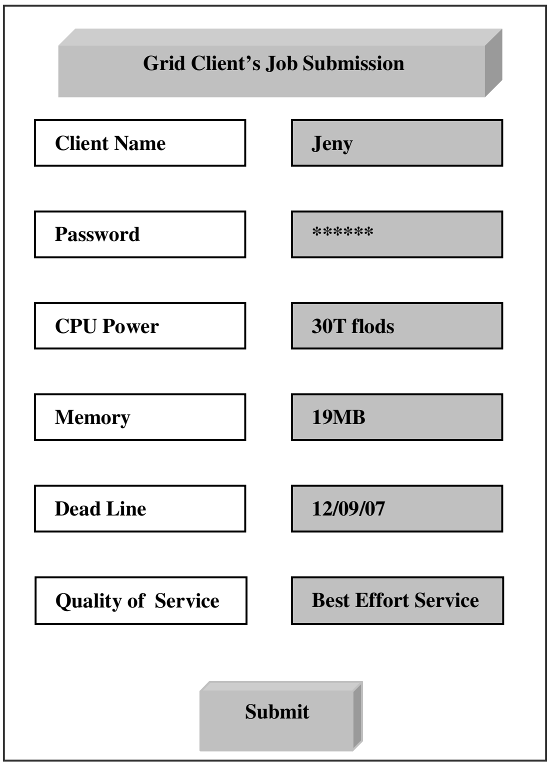


Fig. 1: Job submission blueprint

Literatures have proposed a grid workflow scheduling algorithm in which cost is optimized with the expectation to minimize the makespan. Literatures have also presented a scheduling approach for the economics-driven grids to optimize the cost under the deadline constraint. In fact, a mixed-integer non-linear programming algorithm was introduced to optimize the cost with the consideration of other QoS requirements. As the scale of workflow applications becomes larger and larger, conventional deterministic approaches may fail to give a satisfying solution. Moreover in Grid scheduling problem, for most practical applications, any scheduler delivering good quality planning of jobs would suffice rather than searching for optimality. In fact, in highly dynamic Grid environment, there is no possibility to even define optimality of planning as it is defined in combinatorial optimization. This is due to the fact that Grid schedulers run as long as the Grid system exists and thus the performance is measured not only for particular applications but also in the long run. It is well known that meta-heuristics are able to compute in short time high quality feasible solutions. Therefore, meta-heuristic algorithms have been receiving growing interests due to their powerful global search capability.

From the above exposition we are motivated and in this paper we apply the probabilistic Tabu search algorithm for the generalized Grid Scheduling problem. The basic idea behind the algorithm is to use preprocessing operations to arrive at a probability value for each vertex which roughly corresponds to its probability of being included in an optimal solution, and to use such probability values to shrink the size of the neighborhood of solutions to manageable proportions. We report results from computational experiments that demonstrate the superiority of this method over the generic Tabu search method.

PROBLEM DESCRIPTION

The Super Schedule (SSGA) Grid Architecture described with eight nodes Grid environment example is shown in the Fig 2. This architecture can be utilized for any practical applications for the normal grid environments. The setup is experimented in TIFAC Core in Network Engineering under DST project.

The goal of the SSGA is to find the allocation sequence of workflows on each Grid site. Four major entities are involved in this architecture.

- The grid users submit their request for job completion to the local grid managers.
- All the tasks should be received by the grid managers and the decision for the scheduling is made on deploying the request to the Intra Grid schedulers.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

- The Intra-Grid schedulers have the updated information of the grid resources that are idle during time t. This information is frequently updated. The smaller jobs can be scheduled within their deadlines by the Intra-Grid schedulers in their respective Administrative Domains. Here scheduling is often dynamic.
- For data intensive applications where the jobs are larger it requires the necessity of the resources worldwide. At that moment, there is a necessity of Inter-Grid schedulers which is static often.

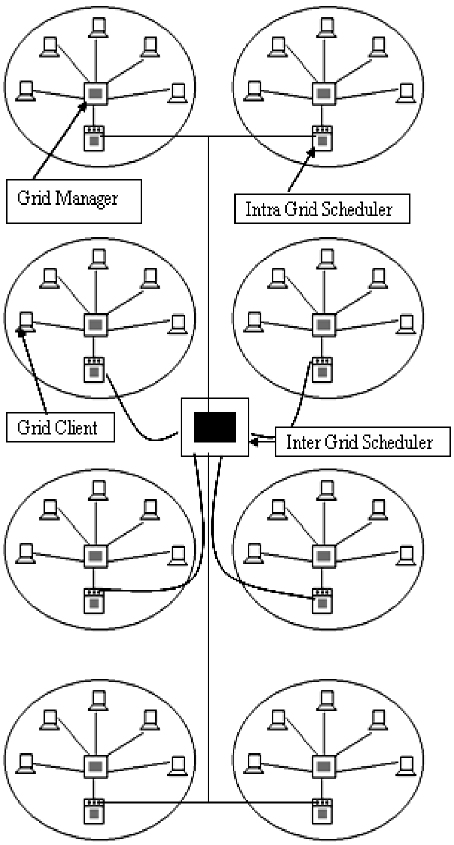


Fig 2: Super Schedule Grid Architecture

The workflow allocation strategy in a Grid environment differs from the traditional ones. The goal of the Inter-Grid Scheduler is to receive the request from different Intra-Grid Schedulers and make an optimistic scheduling such that it accommodates many workflows completing within its deadline. The following DAG workflows and the penalty cost for each workflow are considered for experimental purpose.

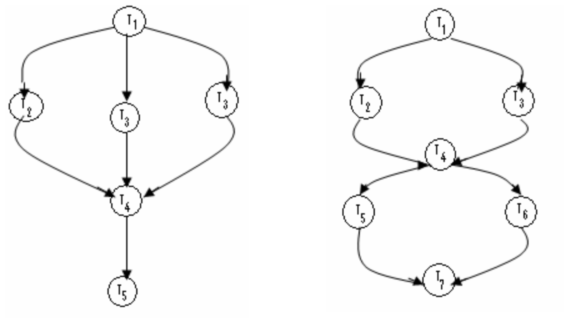


Fig 3: DAG workflow model

The duration for any workflow, penalty cost incurred and the required grid resources are shown in the Table 1.The tasks taken for experiment have their predecessors and successors, such as T1 follow T2 or T2, T3 and are parallel computations once the task T1 is executed.

Table 1: Experimental work flows

Work flow	Tasks/duration	Penalty cost	Grid
W1	T_1(3), T_2(2), T_3(3), T_4(4), T_5(3), T_6(1)	100.00	Any
W2	T_1(2), T_2(4), T_3(3), T_4(1), T_5(3), T_6(2), T_7(3)	50.00	Any
W3	T_1(2), T_2(4), T_3(1), T_4(2), T_5(3), T_6(4), T_7(5)	75.00	Any
W4	T_1(2), T_2(3),T_3(1), T_4(4), T_5(5), T_6(1), T_7(2), T_8(3)	80.00	Any
W5	T_1(1), T_2(3),T_3(2), T_4(4), T_5(5), T_6(6)	100.00	Any
W6	T_1(1), T_2(3),T_3(1), T_4(2), T_5(4), T_6(2)	99.00	Any
W7	T_1(1), T_2(4),T_3(2), T_4(5), T_5(3), T_6(2)	150.00	Any
W8	T_1(1), T_2(3),T_3(2), T_4(3), T_5(1), T_6(1), T_7(2), T_8(3)	100.00	Any
W9	T_1(3), T_2(3),T_3(4), T_4(6), T_5(2), T_6(4), T_7(2), T_8(1),T_9(5)	10.00	Any
W10	T_1(4), T_2(5),T_3(4), T_4(4), T_5(5), T_6(1), T_7(2)	5.00	Any

The Workflow model for W1, W2, W3 are shown in Fig. 3. The FCFS map tasks to the idle Grid sites based on first task arrival to serve first. The EDF algorithm executes the tasks whose absolute deadline is the earliest. Hence it estimates the execution deadline of the individual workflow for any standalone system and schedules such that the workflows that require greater completion time is served first. In EDF the task priorities are not fixed but change depending on the closeness of their absolute deadline.

The settings of the experiment consist of workflows with following assumptions:

- Each workflow received in the Inter-Grid Scheduler consists of a set of Tasks T1, T2, T3 and so on.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

- The task in each workflow is a Directed Acyclic Graph (DAG) model. (Fig. 3.)
- The output from a task can be transferred to other tasks as per the DAG graph model and all jobs are available at time zero.
- At any time a task can be executed only on a Grid site which is reported to the Inter-Grid scheduler as idle via Intra-Grid scheduler.
- There is no pre-emption of tasks or workflows.
- The sequential order of workflow allotment changes.

Here we present a scheduling approach for the wide area problem where in the resources and jobs are dispersed geographically.

PROPOSED METHOD OF PTS

In this study, PTS heuristic to solve scientific workflow scheduling problem in Grid is discussed. The roots of Tabu search go back to the 1970's; it was first presented in its present form by Glover [Glover, 1986]; the basic ideas have also been sketched by Hansen [Hansen 1986]. Additional efforts of formalization are reported in [Glover, 1989], [de Werra & Hertz, 1989], [Glover, 1990]. Many computational experiments have shown that tabu search has now become an established optimization technique which can compete with almost all known techniques and which - by its flexibility - can beat many classical procedures.

The generic TS is a metaheuristic strategy based on neighborhood search with overcoming local optimality. It works in a deterministic way trying to model human memory processes. Memory is implemented by the implicit recording of previously seen solutions, using simple but effective data structures. This approach focuses on the creation of a Tabu list of moves that have been performed recently and are forbidden to be performed for a certain number of iterations, thereby helping to avoid cycling and promoting search in a diversified space. At each iterations, TS moves to the best solution that is not forbidden and thus independent of local optima

The generic TS introduce flexible memory structures articulating strategic restrictions and aspiration levels as a mean for exploiting search spaces. TS have the ability to generate solutions of notably high quality such as to escape from the local minima and to implement an explorative strategy. TS are an iterative procedure for searching a global optimum for discrete combinatorial problem. The philosophy of TS is to avoid entrainment in cycles by forbidding or penalizing moves, which take the solution in the next iteration, to points in the solution space previously visited.

In order to improve the efficiency of the exploration process, one needs to keep track not only of local information (like the current value of the objective function) but also of some information related to the exploration process. This systematic use of memory is an essential feature of Tabu search (TS). While most exploration methods keep in memory essentially the value $f(i^*)$ of the best solution i^* visited so far, TS will also keep information on the itinerary through the last solutions visited. Such information will be used to guide the move from i to the next solution j to be chosen in $N(i)$. The role of the memory will be to restrict the choice to some subset of $N(i)$ by forbidding for instance moves to some neighbor solutions. More precisely, we will notice that the structure of the neighborhood $N(i)$ of a solution i will in fact be variable from iteration to iteration.

The main problem with such a tabu search algorithm is the size of the the neighborhood, for each solution. Thus generic Tabu search is able to execute only a few iterations within reasonable execution times and therefore alleviating the complexity of matching a job to the appropriate resource in the shortest time possible. The Probabilistic Tabu search for Grid scheduling addresses this concern.

SOLUTON CONSTRUCTION

The structure of Probabilistic Tabu search is as shown below. The basic idea is to look at only a subset of the neighborhood of each solution which has the maximum likelihood of containing the best tabu and non-tabu neighbors. The belief is that a large enough set of locally optimal solutions collectively contain predominantly those features that are present in globally optimal solutions and rarely contain features that are absent in globally optimal solutions. In this approach, a pre-defined number of starting solutions are chosen from widely separated regions in the sample space, and used in local search procedures to obtain a set of locally optimal solutions. These locally optimal solutions are then examined to provide an idea about the probability of each solution being included in an optimal solution. Using this idea, the neighborhood of each solution is searched in a probabilistic manner.

General Scheme of PTS: The structure of PTS algorithm is formalized as shown below.

Step 0 (Generating Probabilities): Generate a set of s solutions $S = \{S1,S2, . . . , Ss\}$ using an extension to

local search method to obtain a local optimum. For each solution Si compute the associated probability pi Go to Step 1.

Step 1 (Initialization): Define all solution elements as non-tabu. Choose an initial solution S, set BestSolution ← S, and set Iteration ← 1. Go to Step1.

Step 2 (Termination): If a pre-defined termination condition is satisfied, output BestSolution and exit. Else go to Step 3.

Step 3 (Iteration): Consider each neighbor N of S with a probability of (1−pi)pj where vi = S \ N and vj = N \ S. If vi or vj is marked 'tabu' then N is a tabu neighbor, otherwise it is a 'non-tabu' neighbor. If the best tabu neighbor considered has a cost lower than the cost of BestSolution, go to Step 4, else replace S by the best non-tabu neighbor considered. Mark the solution elements participating in this move (i.e. the vertex that has left the solution, and the vertex that has entered the solution to form the neighbor) as tabu for the next TENURE moves. If this best non-tabu neighbor is better than BestSolution, replace BestSolution with this neighbor. Set Iteration ← Iteration + 1. Go to Step 2.

Step 4 (Aspiration): Replace BestSolution and S with the tabu neighbor of S. Remove the tabu status for all solution elements. Set Iteration ← Iteration + 1. Go to Step 2.

For every solution move in the TS procedure, the neighborhood solution will be evaluated for a Dual Objective Function of minimizing the total penalty cost on choosing the workflow sequence and maximizing the number of workflows completed within deadline (Job Completion Ratio). In our proposed method, the workflows are created based on DAG model and the deadline is fixed to be at 1.5 * Execution time.

RESULTS AND DISCUSSION

The methodology is such that an initial job sequence is selected at random among the set of job sequences and the dual objective function for the solution is defined as a best cost. The obtained solution is recorded as initial step for the Probabilistic Tabu Scheduling mechanism. Later, the set of neighborhood solution of S is generated and again the dual objective function (DOF) is calculated and replaced if necessary finding the best cost among the history record.

The comparative increase in the completion of workflows by PTS dual objective scheduling mechanism considering other algorithms such as FCFS, EDF and TS are shown in Fig 4 and Fig 5.

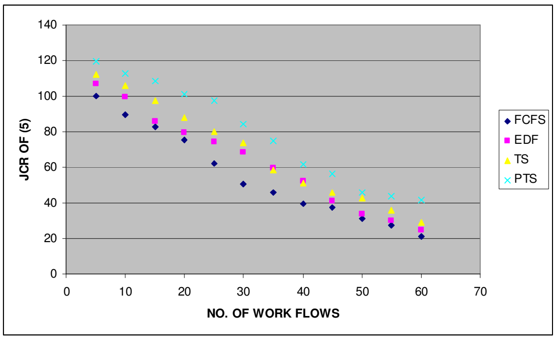


Fig 4: Job completion ratio

It can be analyzed that PTS outperforms TS in the number of workflow completions. In Table 2, the penalty cost incurred by the Inter-Grid scheduler on not completing the job is plotted. As per the methodology PTS succeeds the other scheduling mechanisms in consideration.

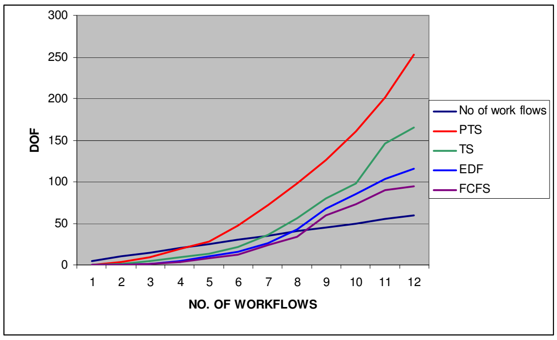


Fig 5: DOF for PTS, TS, EDF and FCFS

Table 2: Penalty cost incurred for the workflow sequence

No of workflows	FCFS	EDF	TS	PTS
5	41.34	29	25	20.88
10	43.75	35.78	29.94	27.63
15	45.67	42.87	33.78	30.84
20	56.45	45.78	40.82	37.62
25	61.45	50.83	51.98	39.652
30	74.55	58.34	59.674	45.67
35	84.3	73.46	68.3	50.64
40	97.55	79.83	74	62.1
45	100.98	87.67	79.56	75.3
50	108.3	97.25	85.65	82.5
55	112.7	106	99.32	89.41
60	119.5	112.3	106.9	100.26

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

CONCLUSION AND FUTURE WORK

In this paper, we have applied probabilistic tabu search algorithm for the Generalized Grid Scheduling problem. In this approach, a pre-defined number of starting solutions are chosen from widely separated regions in the sample space, and used in local search procedures to obtain a set of locally optimal solutions. These locally optimal solutions are then examined to provide an idea about the probability of being included in an optimal solution. Using these ideas, the neighborhood of each solution is searched in a probabilistic manner. Our computational experience shows us that this probabilistic tabu search method outperforms generic tabu search most of the time.

In the near future we plan to combine Probabilistic Tabu search with simulated annealing along with sharing method to increase the efficiency. Similarly the ant colony properties can be included for scalability in the existing algorithm. The procedure can also suitably be modified and applied to any kind of Grid scheduling with different problem environment and optimize any number of objectives concurrently.

REFERENCES

E.H.L. Aarts, P.J.M. van Laarhoven, J.K. Lenstra, and N.L.J. Ulder, "A Computational Study of Local Search Algorithms for Job Shop Scheduling*", ORSA Journal on Computing 6, (1994)118-125.*

I. Foster and C. Kesselman, *The grid: Blueprint for a future computing infrastructure*, San Mateo, CA: Morgan Kaufmann, 1999.

M. Maheswaran, et al., "Dynamic mapping of a class of independent tasks onto heterogeneous computing systems", *Journal of Parallel and Distributed Computing*, Vol. 59, 1999, pp. 107-131.

R. Buyya, D. Abramson, and J. Giddy, "A case for economy grid architecture for service oriented grid computing", *10th Heterogeneous Computing Workshop (HCW' 2001)*, 2001.

I. Foster, C. Kesselman, S. Tuecke, "The Anatomy of the Grid: Enabling Scalable Virtual Organizations", *Intl J. Supercomputer Applications*, 2001.

H. XiaoShan, S. XiaoHe, "QoS guided min-min heuristic for grid task scheduling", *Journal of Comput. Sci. & Technol.*, Vol. 18, No. 4, 2003, pp. 442-451.

Diptesh Gosh, "A Probabilistic Tabu Search algorithm for the Generalized Minimum Spanning Tree Problem" Published in 2003, Indian Institute of Management (Ahmedabad)

A. A. Mandal, et al. "Scheduling strategies for mapping application workflows onto the grid", *in Proceedings of the 14th IEEE International Symposium on High Performance and Distributed Computing (HPDC-14)*, 2005, pp. 125-134.

J. Yu, R. Buyya, and C.K. Tham, "Cost-based scheduling of scientific workflow applications on utility grids", *Proceedings of the 1st International Conference on e-Science and Grid Computing (e-Science' 05)*, pp. 140-147, 2005.

M.M. López, E. Heymann, M.A. Senar, "Analysis of dynamic heuristics for workflow scheduling on grid systems", *in Proceedings of the Fifth International Symposium on Parallel and Distributed Computing (ISPDC'06), IEEE*, 2006.

A. Afzal, J. Darlington, A.S. McGough, "QoS-constrained stochastic workflow scheduling in enterprise and scientific grids", *The 7th IEEE/ACM International Conference on Grid Computing*, 2006, pp. 1-8.

Name:
R. Joshua Samuel Raj

Affiliation:
Assistant Professor / CSE
VV College of engineering.

Brief Biographical History:
2005 -Graduated in 2005 from the Computer Science and Engineering Department from PETEC under Anna University
2007 -Received M.E Degree in Computer Science and Engineering from Jaya College of Engineering under Anna University
2009 Working towards the Ph.D degree in the area of Grid scheduling under Kalasalingam University

Main Works:
Grid computing, Mobile Adhoc Networking, Multicasting and so forth

Name:
V. Vasudevan
Affiliation:
Director, Software Technologies Lab, TIFAC
Core in Network Engineering,
Srivilliputhur, India

Brief Biographical History:
1984- M.Sc in Mathematics and worked for several areas towards Representation Theory
1992 Received his Ph.D. degree in Madurai Kamaraj University
2008- the Project Director for the Software Technologies Group of TIFAC Core in Network Engineering and Head of the Department for Information Technology in Kalasalingam University, Sirivilliputhur, India

Main Works:
Grid computing, Agent Technology, Intrusion Detection system, Multicasting and so forth

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Overclocked Load Scheduling in Large Clustered Reservation Systems

Tania Taami
Islamic Azad University,
Science and Research Branch,
Tehran, Iran
t.taami@srbiau.ac.ir

Amir Masoud Rahmani
Islamic Azad University,
Science and Research Branch,
Tehran, Iran
rahmani@sr.iau.ac.ir

Ahmad Khademzade
Islamic Azad University,
Science and Research Branch,
Tehran, Iran
Zadeh@itrc.ac.ir

Ismail Ataie
Jam Petro. Complex,
Tehran, Iran
ataie.ismail@gmail.com

***Abstract*—Advanced resource reservation has a great role in maintaining QoS of requests. Resource allocation and management to reservation requests for optimal utilization and guarantee of quality of service is challenging effort. When a reservation request for a resource type fails although enough free capacity might be available, there is not any chance for resolving conflicts. Inflexibility of reservation request in support of replacement on time axis, results in rigid resource utilization and even poor QoS of the system. But with the help of new overclocking technologies for doing over-clocking on some current scheduled reservation chunks, new chances emerge to beat these restrictions [1]. Using strict overclocking schema with traditional processors in limited time in cluster of servers, simulation results show QoS of reservations could be improved. This is came through with improvement to utilizing of resources and increasing accepted reservations without any side effects on processing and reliability of computations.**

Keywords-scheduling; overclocking; thermal behaviour; advance reservation; cluster; QoS;

I. Introduction

In center of any collection system should be a scheduler to manage and allocate resources to the clients in appropriate time. Once of most essential resources in any system, either single or orchestrated system is processing unit. Accepting and scheduling requests in appropriate time on appropriate nodes is challenging effort of scheduler. In this paper we concentrate on overclocking computing resource to beat underutilized resources and improving QoS of reservations.

Previously, many efforts have been done for scheduling in clusters or grid systems [2, 6, 7, 8, 9, 10, 11] and also scheduling with over-clocking capabilities in single node systems for real-time (periodic and aperiodic) jobs [1, 5], but no studies about the integration of these yet.

In reliable overclocking, computing resource should be controlled so that does not pass the thermal threshold of equipment [1]. In this paper is introduced simple model of reliable overclocking processors, either overcome complexity of real thermal model of processors that impact any algorithms in real time and either reduce complexity of computation of thermal radiated from processors that also reduce computation time of any stage of algorithm.

Physical architectural model of computing nodes is a cluster of nodes that connected by a shared back bone [12]. Any workload is divided in two subdivisions. In the first division workload is deployed to node or nodes and in the second division workload(s) is started and continued up to its end. After transferring workload(s) to target(s), computation starts and terminates until end of its workload. Two constraints exist on this model: computation capacity of nodes and bandwidth capacity of infrastructure of network.

Using overclocking any reservations or allocation on computing nodes could be relocated, finish times. Computing resources overclocking needs awareness of troubles that might be introduced in reliability of results and on hardware chips. On the other hand, solving thermal equations of node material is costly in real time scheduler [1]. So, for improving the schedulers we need a simple and dependable model to utilize capabilities of resources.

The layout of this letter will be as follows: section II will describe system model, reservation model, overclocking concepts and strict overclocking schema. In section III we will propose an algorithm that combined overclocking and scheduling mechanisms into harmony. We will evaluate the performance of proposed algorithm with the simulation and results in section IV. Finally, in section V we present our conclusions of algorithms and proposed over-clocking schema.

II. Models and Overclocking Concepts

A. System Model

In this paper we choose system models of [12]. At this moment, briefly describe this model.

In this model we have one type of requests: reservation requests. according definition any reservation request R has five parameters: Rc, Rs, Re, n, Rio, where Rc is coming time of reservation request, Rs is start time of reservation, Re is end time of reservation, n is number of processing units that should be served for reservation and Rio is aspect of time is required to transferring reservation request to processing units. In this model requests should be guaranteed to serviced with n processing unit, in interval Rs and Re. Reserves could not coming in system earlier than Rc time

but could out of system earlier than Re time if all of works have been done on computing nodes.

The system model in this paper is considered as a cluster of nodes that connected by a single and shared media backbone, similar to a LAN network. A cluster consist of one coordinator node and n agent nodes A1, A2, ..,An. the coordinator node receives requests, reservations, and possibly plans to schedule request on agent nodes by its scheduler module. In a different way, each agent node also has two major parts: local scheduler and processor frequency controller. The coordinator's scheduler dispatches scheduling timetables and requests that should be ran on node, to agent schedulers. According received timetables local scheduler give control of processing unit to request, the reservation. Figure 1 shows structure of cluster of nodes with a master or coordinator for managing several agent nodes that all connected to single backbone.

According to this model of computation, there are two resource, computing resource and network resource. Based on these two types of resources, there are conflicts on accessing and utilization them. First conflict appears when any two or more request want exclusively access the network media for communicating and deploying workload to destination node. Only one of them could access the network and transfer its data to destination node. Another resource is computing power of the nodes. When a request wants completely access to the node, intended for uses it for processing purposes in some time interval, other requests could not access it until end of processing time of current request on it.

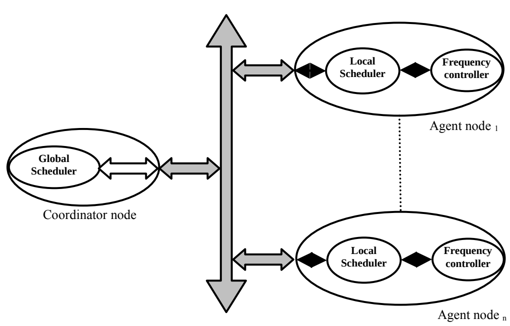


Figure 1. Topology of cluster of nodes with a coordinator and many agent nodes.

B. Thermal Model

Relation between processor speed and thermal behavior of any chip can be approximated by the following equation[1]:

$$T'(t) = \frac{\kappa s^{\alpha}(t)}{C} - \frac{T(t)}{R \cdot C} \qquad (1)$$

Where T(t) is temperature at time t and $s(t)$ is speed of processor at time t. the parameters R and C are the thermal resistance and capacitance of chips, respectively (with fan or any peripheral attached to chip, like heat sink).

The parameter α and κ relate power consumption of processor to its speed. The α parameter has a value of roughly 3.0 [1,3]. For safety of system, processor temperature should not reach to critical point of temperature, due to damaging effects on chip operation.

According to thermal model in the (1), we can derive following (2) for calculating temperature at any point of time[1,3]:

$$T_E = T_F + (T_0 - T_F)e^{-t/\tau} \qquad (2)$$

Where in general $T_F = R\kappa s_F^{\alpha}$ is steady state temperature at overclocking speed of s_F and $T_E = R\kappa s_E^{\alpha}$ is temperature at between times with speed of s_E after elapsed t unit of time, and T_0 is the temperature at lowest level at the start time. Parameter Δ is equal to $R{\cdot}C$ and t is elapsed time of time that temperature was T_0.

By this equation, we can calculate the t value:

$$t = \tau \ln\left(\frac{T_o - T_H}{T_E - T_H}\right) \qquad (3)$$

To avoiding complex and time consuming computations at run time on scheduler, we utilize simple and effective strict overclocking schema. Consequently, in this schema, we exploited three phases in support of CPU frequency scaling, under-clocked phase, normal clocked phase and overclocked phase. In under-clocking phase (i.e. idle mode) frequency of processor is reduced to minimum available value which results in reduced temperature to near the minimum possible value. In the over-clocking phase transiently frequency of processor is increased to maximum value until temperature reach to normal point. Finally in the normal-clocking phase frequency backs to nominal it to continue probably reminded workload of request. Considering the temperature is not above normal, reliability and continuity of computing operations are preserved. Also we cover two working modes in the schema, normal load mode and idle load mode. To reducing temperature more quickly in idle mode we never deploy any workload to the processor that keeps temperature and frequency in lowest limit, i.e. under-clocking phase. We exploit this situation due to expanding succeeding overclocking interval to the maximum possible value. Using the (3) we can calculate t and ratio of under-clocking to over-clocking periods.

III. ALGORITHM

In this section we introduce a scheduling algorithm that uses described strict overclocking schema in situations where conflicts are appeared between current reservation request and previous guaranteed and scheduled requests, reservations parts, is discovered.

As previously described, for overclocking any time period of the processors, we elaborate the three step strict overclocking schema: in first step, node processor get under-clocking frequency with idle workload, in the second, the node get overclocking frequency, and last, the node get normal clocking frequency. Only the timeslots of processor

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

could be overclocked if exists enough timeslot before it that hasn't been allocated to any request.

In following algorithms there are two overclocking approach: other-overclocking and self-overclocking. In other-overclocking approach, timeslot of processor belong to other previous requests, the reservations, is overclocked. But in self-overclocking approach, current request on nodes is overclocked.

The *doReserve* algorithm (Fig. 2) firstly tries to schedule reservation R in cluster of nodes, without over-clocking. If it could not proceed, tries to apply overclocking techniques. The *doReserveWithOverClock* algorithm (Fig. 3) implements a strict overclocking schema that previously has been explained. First it finds eligible nodes; the nodes could be overclocked during period of some scheduled jobs or reservations. If it could schedule by available nodes with normal clocking and overclocking other possible nodes, either self-overclocking or other-overclocking, it proceeds, otherwise it fails. Value of Δ is amount of time that the end of request goes back because of overclocking. The T_{idle} parameter is the required time for period of under-clocking with idle workload.

```
boolean doReserve (R)
if (isFreeIO(R.R_s, (R.R_e- R.R_s)·R.R_io) == false)
  return false;
AvailabeNodes ← findAvailabeNodes(R.R_s, R.Re);
if (#AvailableNodes < R.n)
  return doReserveWithOverClock(R);
else reserveNodes(AvailabeNodes, R.R_s, R.R_e, R.n);
  return true;
```

Figure 2. Top level of reservation algorithm

```
boolean doReserveWithOverClock (R)
// find and set Eligible Allocation scheduled slot of nodes for
// overcloking
EligibleAllocs ← ∅
for i = 1 to n
    Alloc_i=null;
    Δ=min((R.R_e-R.R_s-R.R_io), maxOCTime)*OCRate;
    if (R_id = cpuOverlap(node_i, R.R_s, R.R_e)) != null and
            isFree(node_i, R_id.R_s-T_idle-R_id.R_io, R_id.R_s) and
            (R_id.R_e - Δ) ≤ Rs and
            isFree(node_i, R_id.R_e, R.R_e) )
        TimeInterval_node_i,R ←( R_id.R_s - T_idl, R.R_e-Δ);
        Alloc_i= (node_i, R_id, TimeInterval_node_i,R);
    end if;
    if (Alloc_i !=null)
        eligibleAllocs ← eligibleAllocs + Alloc_i;
end for
AvailabeNodes ← findAvailabeNodes(R.R_s, R.R_e);
if (#EligibleAllocs + #AvailabeNodes ≥ R.n)
  reserveNodes(AvailabeNodes, R, #AvailabeNodes);
//  reserve nodes with overclocking
  for i=1 to R.n - #AvailabeNodes
    RE=EligiblesAlloc_i.R
    Δ = min((RE.R_e-RE.R_s-RE.R_io), maxOCTime)*OCRate;
    EligibleAllocs_i.interval.start -= T_idle;
    EligibleAllocs_i.interval.end -= Δ;
    updateAllocOnNode(EligibleAllocs_i.node,EligibleAllocs_i);
    allocateNode(EligibleAllocs_i.node, R.R_s, R.R_e, R);
  end for;
  return true;
else
// find nodes that have self OverClocking condition for
// Reservation R
  selfOCNodes ← ∅
  for i=1 to n
    if (isFree(node_i, R.R_s- T_idle, R.R_e-(Δ))
      selfOCNodes +=node_i;
  end for
  if (#EligibleAllocs+ #selfOCNodes+#AvailabeNodes
                                                  ≥ R.n)
   reserveNodes(AvailabeNodes, R, # AvailabeNodes);
// reserve nodes for R reservation with overclocking other
// scheduled requests
   for i=1 to R.n - #AvailabeNodes
     RE=EligiblesAlloc_i.R
     Δ = min((RE.R_e-RE.R_s-RE.R_io),maxOCTime)*OCRate;
     EligibleAllocs_i.interval.start -= T_idle;
     EligibleAllocs_i.interval.end -= Δ;
     updateAllocOnNode(EligibleAllocs_i.node,
                                        EligibleAllocs_i);
     allocateNode(EligibleAllocs_i.node, R.R_s, R.R_e, R);
   end for;
// reserve nodes for R Reservation with Overclocking R itself
    for  i=1 to  R.n- (#EligibleAllocs+ #AvailabeNodes)
      Δ = min((R.R_e-R.R_s-R.R_io), maxOCTime)*OCRate;
      allocStartTime = R.R_s - T_idle;
      allocEndTime= R.R_e - Δ;
      allocateNode(node_i, allocStartTime, allocEndTime, R);
    end for;
    return true;
  end if;
 end if;
return  false;
```

Figure 3. Strict over-clocking schedular algorithm

Overclocking schema could be applied on start time of computation until end time of it. That is to say, overclocking couldn't be applied on communication part of request because communication time of any request depended to network specification of cluster (i.e. bandwidth) and could not be altered or increased without changing physical characteristics of underlying network's components.

IV. Performance Evaluation

For analysis of mentioned strict overclocking schema, we simulate a cluster of nodes with varying processing nodes and reservation requests. In all simulations, maximum number of requested nodes by any reservation request is number of nodes in cluster. The reservation requests deploy

its workload to the nodes by using multicasting approach, aimed to maximize bandwidth utilization.

For simulating previous algorithms, we use following parameters: Arrival time of reservation requests have Poisson distribution with average of 50 unit of time. Initially we consider length of requests be near to overclocking period, i.e. in interval of [40 .. 50], with uniform distribution that is named Δ. This value of Δ is nearly double of overclocking time length. Secondly we studied multiples of the Δ in system utilization and acceptance ratio of system. For computing fractions of idle time to overclocking time, we used Dell Latitude D810 with Centrino processor and (3). Based in this provision, this ratio calculated as 3 to 2, 3 units of time for idle time and 2 units of time for overclocking time. As mentioned previously, number of requested nodes in each reservation is in [1 .. number of nodes] interval, i.e. with increasing number of nodes, request of nodes for each reservation will rise. Total simulation time, 11 hours was considered. Yield of overclocking than normal operation of processor is 0.5 (the OCRate in the algorithm 2). Also communication time ration or the Rio is 0.1 of total workload. Although advance reservation is used for guarantee of QoS of mixed typical job and reservation for reservation request, in this model we detach start of service and start of request for adapting with future advance reservation models, and simulation purposes (FIFO model).

Results (Fig. 2) show that using strictly overclocking schema improves utilization of resources and acceptance ratio of reservation request in scalable form.

Overall, because of multi node reservation request that is responded through dynamic and elasticity of overclocking, that impacts and results in more utilization in overclocked schema than normal clocking schema, despite of reducing and convergence of overclocked and normal schema together.

Fig. 3 in comparison with Fig. 2 proves that increasing number of nodes have not any impact on improving utilization and acceptance ratio similar to normal clocking.

In other way, with increasing average length of reservation workloads, overall overclocked utilization improvement with respect to normal clocking, will be increased. The reason is that, with increasing the workload, side effects of idle time slice that happened before any overclocking part of workload, is decreased. But with growing number of requests at the constant workload rate, this gain is starting to be decreased, because side effects of underutilized idle times before any overclocked time slices will be raised.

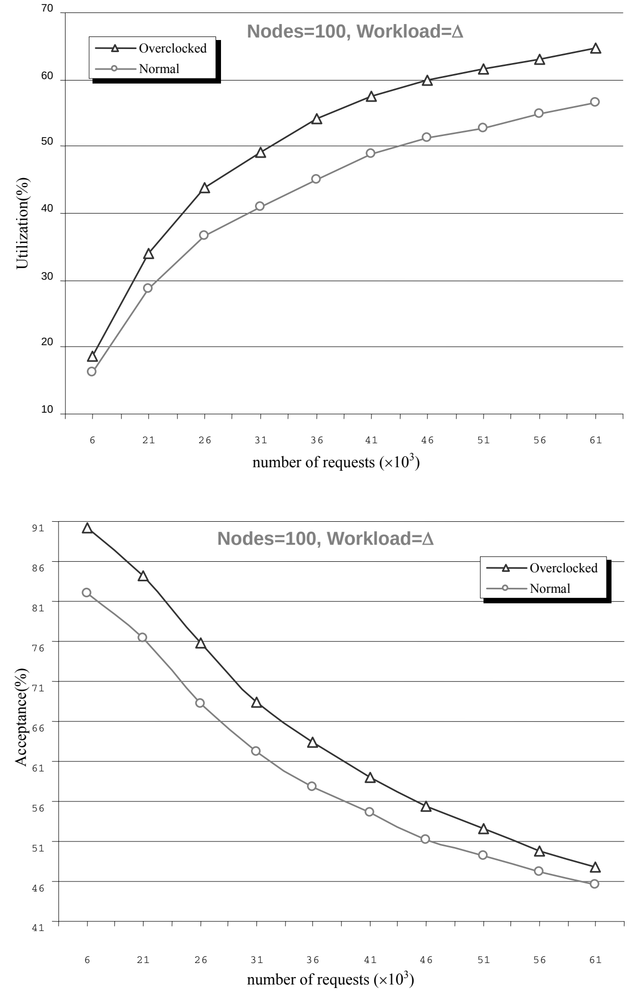
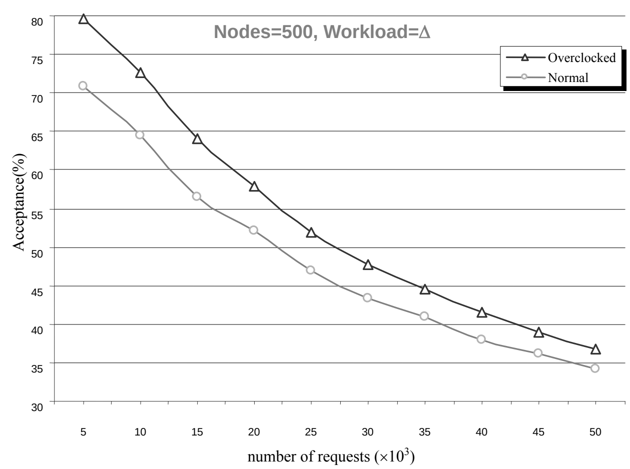


Figure 4. Acceptance and utilization in 100 nodes.



http://sites.google.com/site/ijcsis/
ISSN 1947-5500

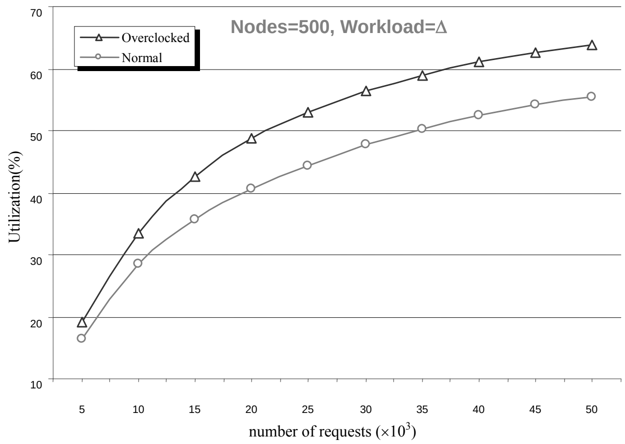
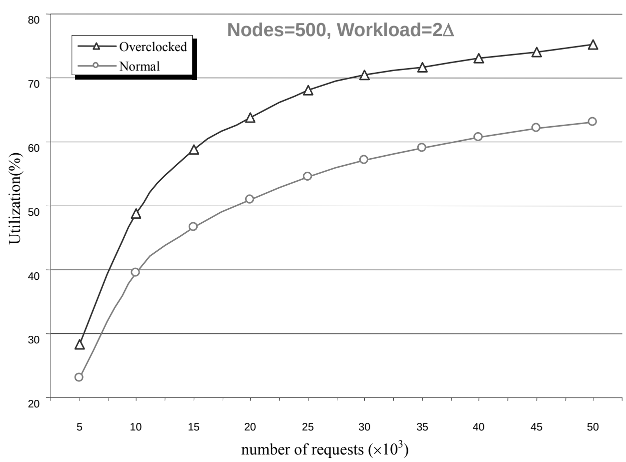
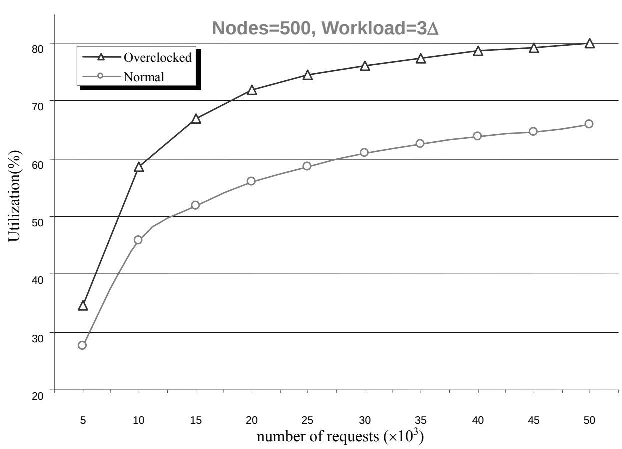
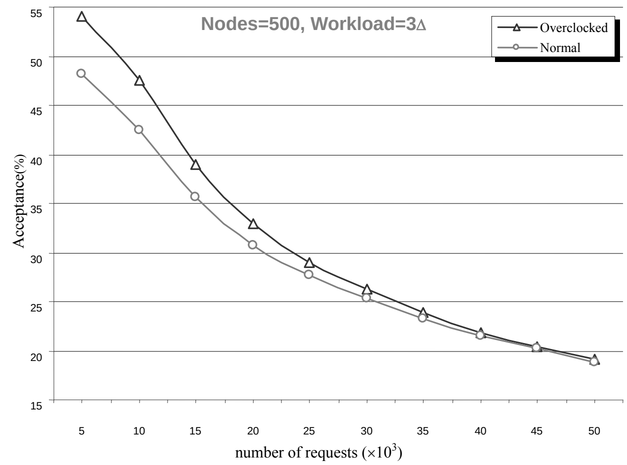


Figure 5. Acceptance and utilization in 500 nodes

In all cases, normal and over-clocked schema, increasing average length of reservations will cause drop of acceptance ratio of reservation requests. Coming out such results is obvious; because of increasing length of reservations, the probability of facing of them with each other will increase simultaneously.







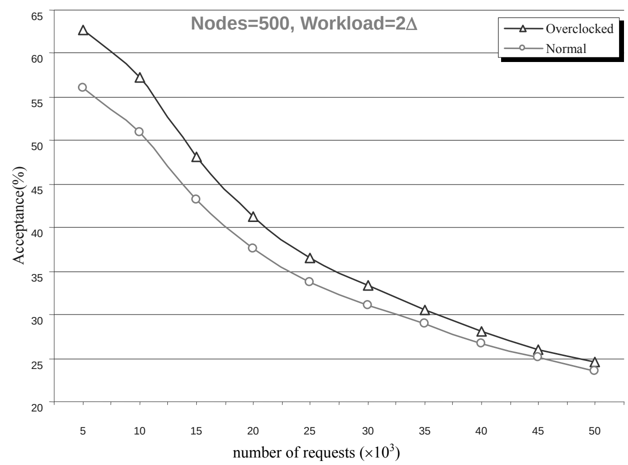
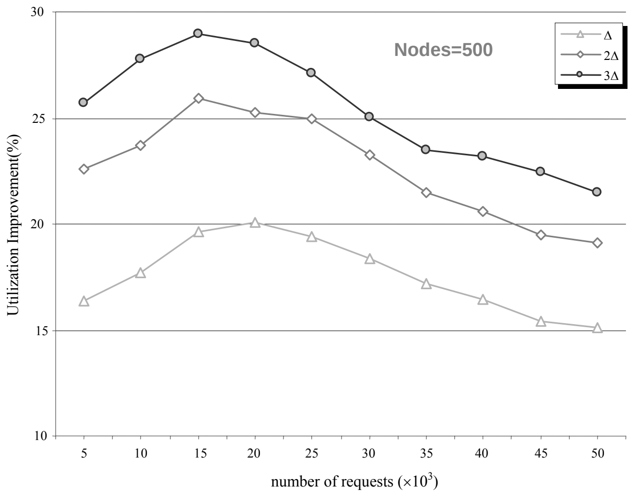


Figure 6. Acceptance and utilization in 500 nodes with workload of 2Δ and 3Δ.

With increasing the workload length of reservations absolutely, both normal and overclocked schemas quickly improve more than before until to reach saturation point. At this point, increasing number of requests, the overclocking has no other influences. Fig. 5 with Fig. 6 shows this matter. Based on default value of Δ, 2Δ, and 3Δ, Fig. 7 graphs show that increasing average workload of requests, peak point of improvement is shifted to left, i.e. towards to less reservation request numbers. This means, with increasing workload, collision between end time of requests and required idle time intervals before overclocking time of processor, will happen sooner.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500



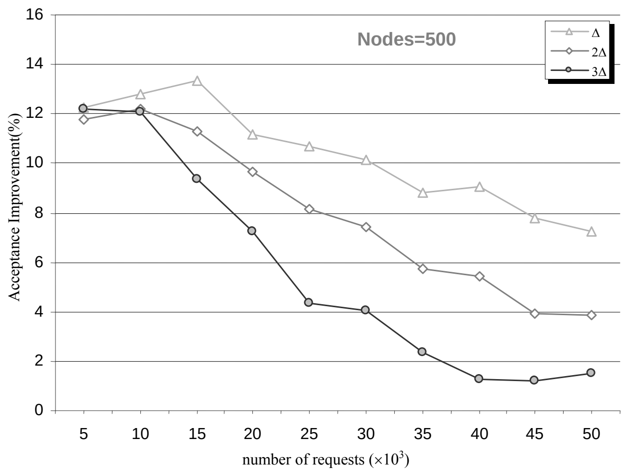


Figure 7. Acceptance and utilization improvement in 500 nodes with workload of Δ, 2Δ and 3Δ.

V. Conclusions

Study of results shows that by means of the proposed strict overclocking schema in controlled boundary, utilization absolutely increases than normal clocking. Also, acceptance rate of system with limited conditions increase. In addition, as temperature of processing nodes could not reach to critical point, reliability of computation is preserved. With preserving power of processor, economical and commercial aspect of power consumption remains.

Expanding networks and resources, we can use this schema in larger grid networks than clusters. Since resources exclusively are provided to requests, this model and algorithms is very good for private grids that total resources available for commercial purposes.

References

[1] Y. Ahn, R. Bettati, “Transient Overclocking for Aperiodic Task Execution in Hard Real-Time Systems”, Euromicro Conference on Real-Time Systems (ECRTS '08), Prague, p. 102, 2008.

[2] D. G. Feitelson, “Scheduling parallel jobs on clusters”, High Performance Cluster Computing, vol. 1, Architectures and Systems, pp. 519–533, 1999.

[3] N. Bansal and K. Pruhs, “Speed scaling to manage temperature”, in Symposium on Theoretical Aspects of Computer Science, 2005.

[4] N. Bansal, T. Kimbrel, and K. Pruhs, “Dynamic speed scaling to manage energy and temperature”, IEEE Syposium on Foundations of Computer Science, 2004.

[5] S. Wang, R. Bettati, “Reactive Speed Control in Temperature-Constrained Real-Time Systems”, Proceedings of the 18th Euromicro Conference on Real-Time Systems (ECRTS 06), Dresden, Germany, pp. 161-170, July 2006.

[6] L. Eyraud-dubois , G. Mounié , D. Trystram, “Analysis of Scheduling Algorithms with Reservations”, Proceedings of the 21st IEEE International Parallel and Distributed Processing Symposium, USA, 2007.

[7] J. Blazewicz, P. Dell’Olmo, M. Drozdowski, P. Maczka, “Scheduling multiprocessor tasks on parallel processors with limited availability”, European Journal of Operational Research, vol. 149, pp. 377–389, 2003.

[8] J. Blazewicz, M. Machowiak, J. Weglarz, M. Kovalyov, D. Trystram, “Schedulingmalleable tasks on parallel processors to minimize the makespan”. Annals of Operations Research, vol. 129, pp. 65–80, 2004.

[9] K. Jansen. “Scheduling malleable parallel tasks: An asymptotic fully polynomial time approximation scheme”, Algorithmica, vol. 39, pp. 59–81, 2004.

[10] O.H. Kwon, K.Y. Chwa, “Scheduling parallel tasks with individual deadlines”, 6th International Symposium on Algorithms and Computation, Springer-Verlag, vol. 215, pp. 198–207, 1995.

[11] V. Subramani, R. Kettimuthu, S. Srinivasan, P. Sadayappan, “Distributed Job Scheduling on Computational Grids Using Multiple Simultaneous Requests”, IEEE Computer Society, p. 359, 2002.

[12] A. Mamat, Y. Lu, J. Deogun, S. Goddard, “Real-Time Divisible Load Scheduling with Advance Reservation”, Euromicro Conference on Real-Time Systems (ECRTS '08), Prague, pp. 37-46, 2008.

Acknowledgment

This work was supported by Iran Telecommunication Research Center (ITRC).

Skew Correction and Noise Reduction for Automatic Gridding of Microarray Images

Manjunath S S,
Asistant Professor, Dept of Computer Science
Dayananda Sagar College of Engineering,
Bangalore. India
Email: mnj_ss2002@ yahoo.co.in

Dr. Lalitha Rangarajan
Reader, Dept of Studies in Computer Science
University of Mysore, India
Email: lali85arun@yahoo.co.in

Abstract-
Complementary DNA (cDNA) microarrays are a powerful high throughput technology developed in the last decade allowing researchers to analyze the behavior and interaction of thousands of genes simultaneously. The large amount of information provided by microarray images requires automatic techniques for efficient processing of microarray images to arrive at accurate biological conclusion. Most of the methods discussed in the literature need different levels of human intervention, which inevitably reduces the efficiency and reproducibility of the entire automation process. In this paper a novel approach for automatic gridding of skewed and noisy microarray images is presented. The microrarray image is skew corrected, noise removed using adaptive thresholds computed on various segments, spatial topology of spots detected, gridding performed and finally grids are refined. Experiments conducted on selected microarray images (skewed and noisy) of Stanford and UNC databases are encouraging

Keywords: Microarray, Gridding, Adaptive threshold, Spatial topology, Grid refinement, Skewed images, Noisy images.

1. Introduction

DNA microarray technology has a large impact in many application areas, such as diagnosis of human diseases and treatments (determination of risk factors, monitoring disease stage and treatment progress, etc.), agricultural development (plant biotechnology), and quantification of genetically modified organisms, drug discovery, and design. In cDNA microarrays, a set of genetic DNA probes (from several hundreds to some thousands) are *spotted* on a slide. Two populations of mRNA, tagged with fluorescent dyes, are then hybridized with the slide spots, and finally the slide is read with a scanner. The outlined process produces two images, one for each mRNA population, each of which varies in intensity according to the level of hybridization represented as the quantity of fluorescent dye contained in each spot.

Microarray image processing consists of the following sequence of three main tasks 1. Gridding, separation of spots by assignment of image coordinates to the spots. 2. Segmentation, separation between the foreground and background pixels and 3. Intensity extraction, computation of the average foreground and background intensities for each spot of the array.

Gridding is an important task that is to be performed as accurately as possible, since it affects the subsequent tasks of segmentation, intensity extraction and finally the conclusions derived out of the whole analysis. The available gridding software packages Scanlyze [1], Dappel [2], Image Gene[3], Genepix and SpotFinder[4] require human intervention in order to specify input parameters as well as to adjust properly the location of the grid lines. The template based approach is most prevalent in the existing packages which require specfication of parameters such as spot size, spot spacing and space location. Some software products already incorporate an automatic refinement search for grid location, given size and spacing of spots [2,3]. Irregular grids cannot be found with the template based approach unless the template is manually adjusted to fit predefined distortions [3]. Automating this part of the process is essential because it reduces error in grid that may arise due to inaccurate specification.

The problem of automatic gridding is complicated because microarray images are usually highly contaminated with the noise and artifacts of the wet lab processes. Rotations, misalignment and local deformations of the ideal rectangular grid can often occur. There is a high need for automated methods for microarray gridding which are robust and flexible at the same time.

Some efforts on automatic microarray gridding have been reported in literature. However most of them impose different kinds of restrictions and are based on stringent assumptions. For example, the approaches in Jain et al.[5] and Yan et al.[6] requires that the grid rows and columns are strictly parallel to the x and y image axes. Other approaches, such as described by Carstensen et al.[7] and Katzer [8], rely on the Bayesian paradigm to deal with

uncertainty and noise. Some well-known approaches to gridding microarray images are based on axis projections (Deng et al.[9]), or on morphological filtering (Yan et al.[10]). Both of them require user intervention in order to manually adjust the grid location. The Hill-Climbing approach for automatic gridding (Rueda et al.[11]) can perform gridding properly only if misalignments and rotations of the ideal grid are not present. Markov random field (MRF) (Antoniol et al.[12]) and graph-based grid approaches (Jung et al.[13]) have been used to perform gridding. A drawback of these approaches is that they require input parameters. A variety of different methodologies have been proposed with the intension to solve rotation and misalignment problems. Bajcsy [14] has suggested an exhaustive search of all the expected rotation angles, where as Steinfath [15] has estimated the rotation angle. A drawback of these approaches is that, it introduces pixel distortions when the rotation angle is small. Brandle et al.[16] utilized the discrete Radon transformation to estimate the angle of rotation. As it is computationally expensive, the process is accelerated by constraining the range of rotation angles. Ho et al.[17] expressed the gridding process as an optimization problem based on the Jacobi iteration. However, this method is efficient only when the grids are smoothly distorted. Giuiliano et al.[18] recommended a gridding procedure based on stochastic search algorithms. Although it deals with rotations effectively, it requires manual intervention in order to define the radius of the spots.

In spite of the potential importance of gridding approaches in microarray image analysis, the existing gridding methods pay little attention to pre-processing of noisy microarray images and focus mainly on spot localization and spot segmentation. The aim of the present study is to propose a method that can deal with rotations, misalignments, and local deformations of the ideal rectangular grid. It is also noise-resistant and it is efficient even under adverse conditions such as the appearance of various spot sizes or the absence of spots.

Figure. 1 Shows a block diagram which describes the salient stages of the proposed approach for automatic gridding of microarray images.

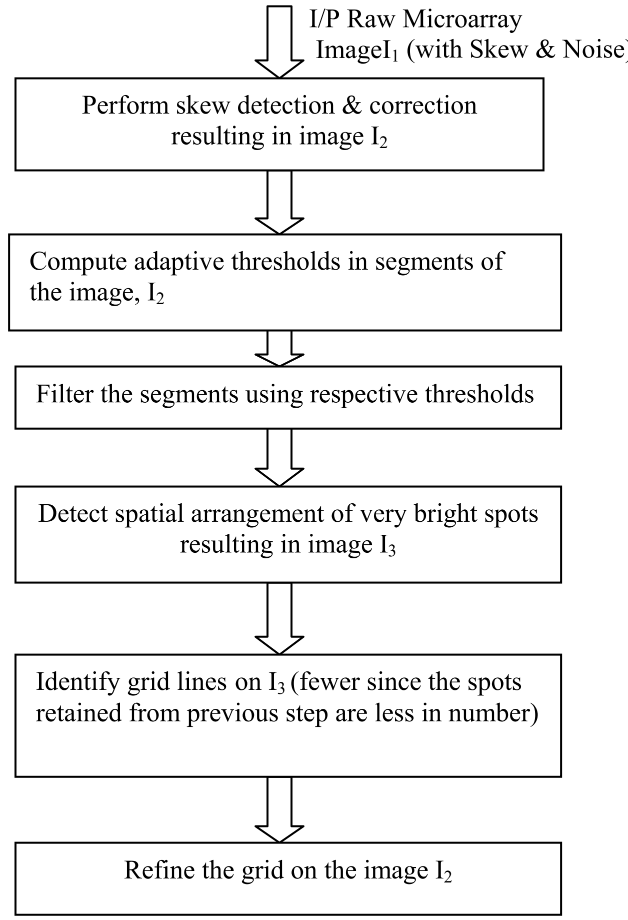


Figure 1. Stages of automatic gridding of noisy Microarray image

The organization of rest of the paper is as follows: In section 2, preprocessing techniques such as skew detection and correction, filtering through respective thresholds are described. In section 3 automatic gridding process which consists of detection of rmin & rmax, detection of cmin & cmax and gridding method are described. Section 4 describes grid refinement algorithm. Section 5 highlights the results of extensive experimentation conducted on some benchmark images. Finally conclusion is discussed in section 6.

2. Skew Detection and Correction

This section describes the first stage of microarray image gridding, that is skew detection and correction.

2.1. Skew Detection

First step in this process is to convert the rgb image to gray scale image. Figure 2. shows the computation of the parameters topx, topy, leftx, lefty, xmid and ymid which are required to find the skew angle. Scan the gray scale image rowwise. The very first pixel in the image is assigned the coordinate address (topx, topy). Scan the gray scale image columnwise. The very first pixel in the image is assigned the coordinate address (leftx, lefty). xmid is the mid value of columns and ymid is the mid value of rows.

.

If topx < xmid and lefty > ymid the skew is clock wise (Figure.2)

If topx > xmid and lefty < ymid the skew is anticlock wise (Figure.3)

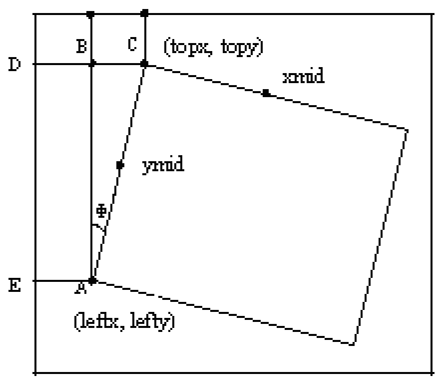


Figure 2. Parameters for clockwise skew detection

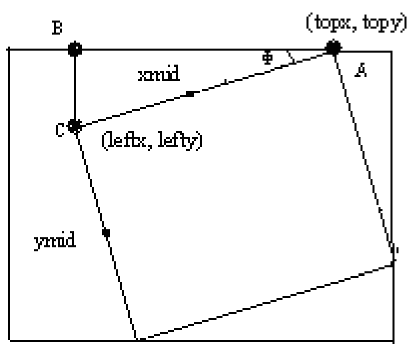


Figure 3. Parameters for anticlockwise skew detection

The clockwise skew angle can be found using the formula

$$\Phi = \text{atan } (topx - leftx) / (lefty - topy)$$

The anticlockwise skew angle can be found using the formula

$$\Phi = \text{atan } (lefty - topy) / (topx - leftx)$$

2.1.2 Skew Correction

The new coordinate address xx and yy are computed as given below.

Skew correction for clockwise tilt: It is required to perform rotation about (leftx,lefty) by Φ in anticlockwise direction .

$$xx = leftx + (x\text{-}leftx) * \cos(\Phi) - (y\text{-}lefty) * \sin(\Phi)$$

$$yy = lefty + (x\text{-}leftx) * \sin(\Phi) + (y\text{-}lefty) * \cos(\Phi)$$

Skew correction for anticlockwise:

$$xx = topx + (x\text{-}topx) * \cos(\Phi) + (y\text{-}topy) * \sin(\Phi)$$

$$yy = topy + (y\text{-}topy) * \cos(\Phi) + (topx\text{-}x) * \sin(\Phi)$$

where
x varies from 1 to number of columns
y varies from 1 to number of rows

The min xx and min yy are computed and translated this to (0,0).The new image L^1 with the coordinate address
xx^1=xx-minxx
yy^1=yy-minyy is the skew corrected image.

Figures 4 and 5 (Image ID: 62919) shows the clockwise skewed image and skew corrected image.

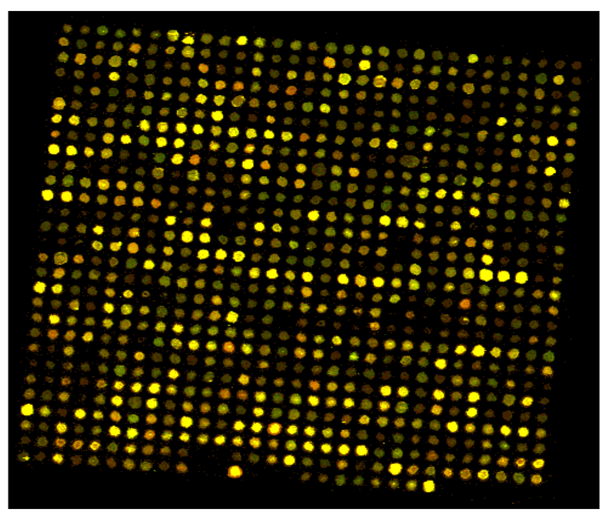


Figure 4. Clockwise Skewed Image ID: 62919

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

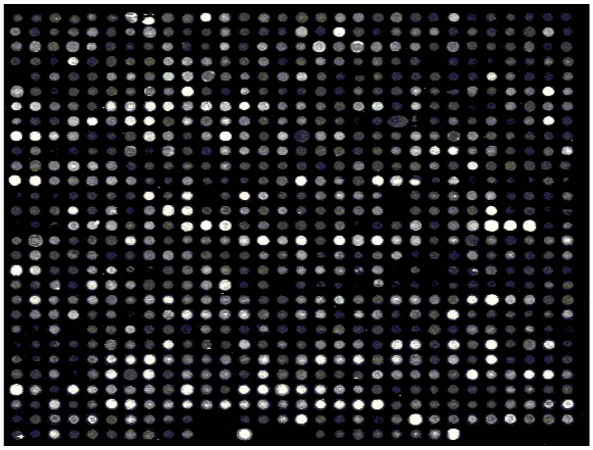


Figure 5. Skew Corrected Image ID: 62919

TABLE 1: ESTIMATED ANGLE (Φ) AND EXECUTION TIME (T_S) OF THE PROPOSED SKEW CORRECTION TECHNIQUE.

Image ID	Estimated angle in degrees (Φ)	Execution time in seconds (τ_s)
1c7b060rex2	2.0651	13.14
1c4bo64rex2	2.5323	12.12
62919	4.8253	15.15
40031	3.7653	13.54

2.2 Adaptive Threshold and Filtering

Filtering is performed in 2 steps which are described in the 2 subsections below. Thresholds on spot size are first computed on segments of the image. Insignificant spots are filtered using these thresholds.

The skew corrected binary image is divided into n segments. Number of segments can be increased depending on the level of noise. The subgrid is divided into 4 segments in the proposed approach as follows

1st segment Rows=0 to r/2 Columns=0 to c/2	2nd segment Rows= 0 to r/2 Columns= c/2+1 to c
3rd segment Rows=r/2+1 to r Columns= 0 to c/2	4th segment Rows= r/2 +1to r Columns= c/2+1 to c

where r is the number of rows and c is number of columns of skew corrected image.

For each segment, the numbers of connected components are computed. The thresholds on spot size in each segment are calculated using the equation below.

$$T(i) = \frac{\text{(Number of pixels in i th segment)}}{\text{(Total number of connected Components)}}$$

Where i ranges from 1 to 4.
For example in the image (ID: 40031) Fig. 6 the number of bright pixels in the four segments are 2462, 1572, 1353, 1065. Total number of connected components is 146. The thresholds are 2462/146=17, 1572/146= 11, 1353/146= 10, 1065/146=9
The results of the proposed filtering process in removing the insignificant spots using the threshold value and execution time (τ_f) are reported in Table 2.

TABLE 2. ESTIMATED THRESHOLD VALUE (A_T) AND EXECUTION TIME (T_F) OF THE PROPOSED FILTERING PROCESS.

Noisy Image ID	Thresholds on a subarray	#spots in the subarray	Execution time(τ_f) in seconds
40031 (Stanford)	T1=17 T2=11 T3=10 T4=09	146	10
44004 (TBDB)	T1=14 T2=12 T3=12 T4=10	777	15
17931 (Stanford)	T1=28 T2=23 T3=47 T4=41	721	13
39119 (TBDB)	T1=20 T2=18 T3=15 T4=15	562	8

Execution time for the filtering process is proportional to number of spots in a noisy microarray image. Adaptive thresholds obtained in the previous step are used to filter insignificant noisy spots in the segments. If the number of pixels in a component are less than threshold value ($T_{(i)}$) in each segment, then remove the spot (insignificant spot) by setting intensity zero to all pixels in that component. The idea behind using adaptive threshold is, if in a subarray should few successive columns or rows have tiny spots filtering using global threshold will eliminate all these spots. This results in sparse grid lines.

Shown in Figure. 6 is a noisy microarray image. Fig. 7 is the noise free filtered image.
.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

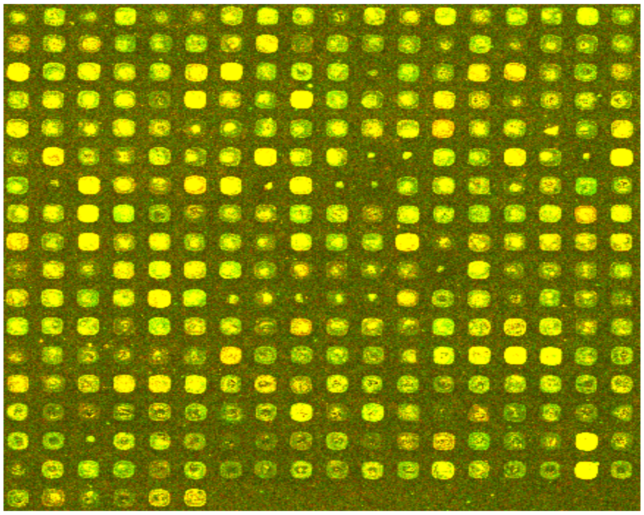


Figure 6. Noisy Microarray Image ID : 40031

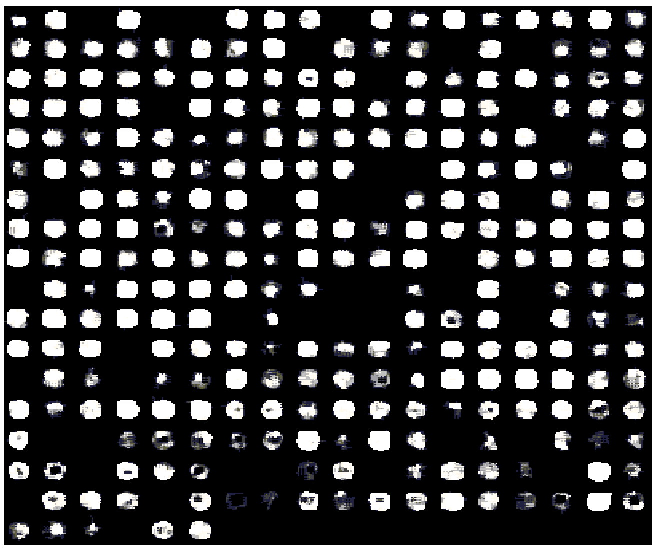


Figure 7. Filtered Microarray Image ID : 40031

3. Automatic Gridding Process

Automatic gridding is performed in 3 steps which are described in the 3 subsections below.

3.1 Determination of position of grid lines

For each connected component in the filtered image, rmin, rmax, cmin, cmax are determined as shown in Fig 8. Sorted arrays of rmin values (similarly rmax, cmin, cmax values) are found. Array of successive differences of rmin array called diff_rmin also for rmax, cmin, cmax (diff_rmax, diff_rmin, diff_cmin, diff_cmax) is found. Key portions of rmin, rmax and diff_rmin, diff_rmax are shown below. All computations done on image ID (62919).

rmin:

60	213	9	110	297	77	246	76	315	9	246
315	43	110	128	9	25	212	333	366	109	60
383	93	129	146	313	76	212	145	128	213	25
211	112	399	163	351	24	177	76	399	297	92.

rmax:

72	226	25	123	310	89	259	90	327	21	258
327	55	121	142	20	38	224	341	377	122	71
394	105	140	156	324	89	224	158	140	222	37
224	124	410	172	360	38	191	87	410	306	104

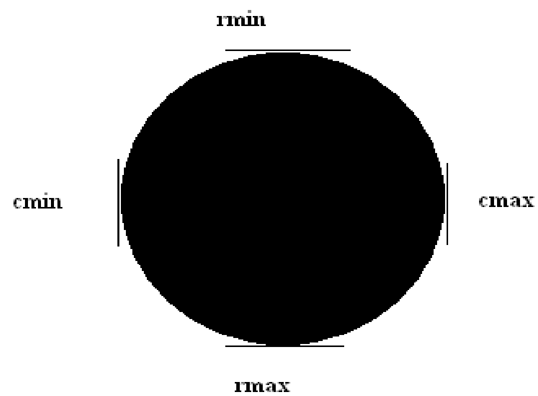


Figure 8. Representation of rmin, rmax, cmin & cmax of a spot.

The steps below describe determination of horizontal grid lines.

1) The rmin array is sorted in ascending order

sorted rmin:

9	9	9	24	25	25	43	60	60	76	76
76	77	92	93	109	110	110	112	128	128	129
145	146	163	177	211	212	212	213	213	246	246
297	297	313	315	315	333	351	366	383	399	399

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

2) The differences of successive rmin values in the sorted rmin array are calculated.

diffrmin:

0	0	15	1	0	18	17	0	16	0	0
1	15	1	16	1	0	2	16	0	1	16
1	17	14	34	1	0	1	0	33	0	51
0	16	2	0	18	18	15	17	16	0	

3) Sudden change in the difference in rmin values indicate the end of previous row of spots and beginning of next row of spots.

4) Observe the sudden change from 0 to 15, at position 3 in diff_rmin array. The third element of rmin array is 9. Hence examination of rmin diff_rmin suggests a grid line at row 9. Similarly it is understood that successive values of grid rmin.

grid rmin:

9	25	43	60	77	93	112	129	146	163	177
194	213	230	246	263	280	297	315	333	351	366
383	399									

Similarly grid_rmax is determined. Shown below are sorted_rmax, diff_rmax, grid_rmax values.

sorted rmax:

20	21	25	37	38	38	55	71	72	87	89
89	90	104	105	121	122	123	124	140	140	142
156	158	172	191	222	224	224	224	226	258	259
306	310	324	327	327	341	360	377	394	410	410

diff_rmax:

1	4	12	1	0	17	16	1	15	2	0	1
	14	1	16	1	1	1	16	0	2	14	2
	14	19	31	2	0	0	2	32	1	47	4
	14	3	0	14	19	17	17	16	0		

grid_ramx:

25	38	55	72	90	105	124	142	158	172	191
208	226	243	259	276	293	310	327	341	360	377
394										

Finally, positions of horizontal gridlines are determined by finding average of rows suggested by grid_rmin and grid_rmax contents. Thus horizontal gridlines are placed at rows 9, 25 (25+25/2), 41 (38+43/2), 57 (55+60/2)…etc.

In a similar manner vertical gridlines are positioned using sorted_cmin, diff_cmin, grid_cmin, sorted_cmax diff_cmax, grid_cmax.

4. Grid Refinement Algorithm

The algorithm described in section before, will determine all grid lines as long as a spot exists on each row and each column of the filtered image. However there may be images where no spots are present in several consecutive rows or columns. In these images, there will be irregular spacing between gridlines. In such cases the refinement algorithm suggested , can be used to draw additional / missing grid lines. Grid refinement process is called to check whether all the gridlines have been drawn. If the differences in the positions of successive rows (i, i+1) is greater than the average of previous spacing of rows(avgrowspace), then the algorithm will draw horizontal lines at every successive avgrowspace beginning from the previously drawn horizontal line, until i+1 or end of rows. Similar procedure is repeated while drawing vertical lines.

Figure 11 shows additional gridlines placed on gridded image of figure 9.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

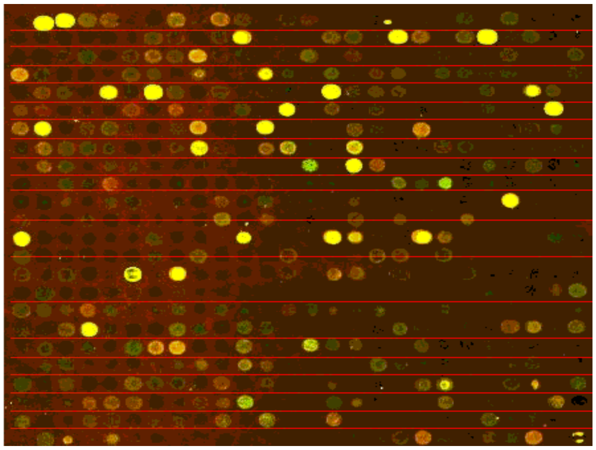


Figure 9. Filtered image with horizontal sparse grid lines.

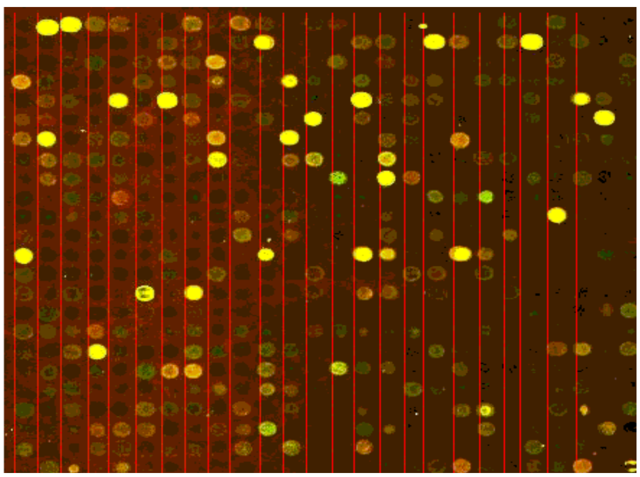


Figure 10. Filtered image with vertical sparse grid lines

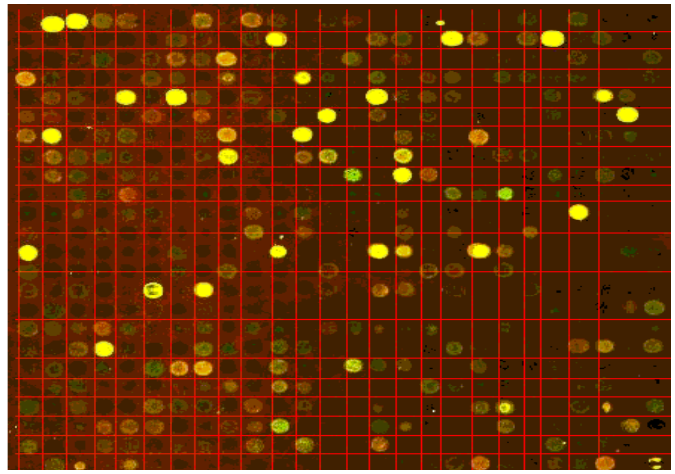


Figure 11 .Gridding of noisy microarray image before refinement process

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

4. Experimental Results

In this section the performance of the proposed approach is evaluated on skewed and noisy microarray images from SMD (Stanford Microarray Database) and UNC (University of North California Microarray Database).The images are available for free download from https://genome.unc.edu.
The algorithm was executed on Pentium 4 proceesor with 2 GB RAM. The results are summarized in the table 3

Table 3 shows comparison of proposed method with projection profile algorithm and standard deviation algorithm to perform gridding. The comparison was performed on 10 sets of microarray images and it is evident that proposed method performs better than other existing approaches. Expected number of rows and columns are inferred by the number of connected components across each row and column.

.

TABLE 3: COMPARISON OF PERFORMANCE OF PROPOSED APPROACH AND OTHER APPROACHES

Method	**Image ID**	**Expected Number of Rows**	**Expected Number of Columns**	**Number of Rows obtained**	**Number of Columns obtained**	**Total Error (%)**
Grid Using Standard Deviation	62919	29	30	27	27	8.474576
	22593	17	15	21	15	12.5
	37993	29	29	27	29	3.448276
	34212	20	21	21	21	5.882353
	34217	18	23	18	23	0
	34143	22	23	22	21	4.444444
	34134	23	23	23	22	2.173913
	52694	28	29	23	28	10.52632
	57852	27	29	25	28	5.357143
	66357	28	29	26	29	3.508772
Grid Using Projection Profile	62919	29	30	27	29	5.084746
	22593	17	15	20	15	9.375
	37993	29	29	26	26	10.34483
	34212	20	21	20	21	11.76471
	34217	18	23	21	24	9.756098
	34143	22	23	24	23	4.444444
	34134	23	23	23	21	4.347826
	52694	28	29	26	29	3.508772
	57852	27	29	25	29	3.571429
	66357	28	29	27	29	1.754386
Grid Using Proposed Approach	62919(SMD)	29	30	29	30	0
	22593(SMD)	17	15	17	15	0
	37993(UNC)	29	29	29	29	0
	34212(UNC)	20	21	22	25	0
	34217(UNC)	18	23	18	23	0
	34143(UNC)	22	23	24	23	4.444444
	34134(UNC)	23	23	23	23	0
	52694(SMD)	28	29	28	29	0
	57852(SMD)	27	29	27	29	0
	66357(SMD)	28	29	28	29	0

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

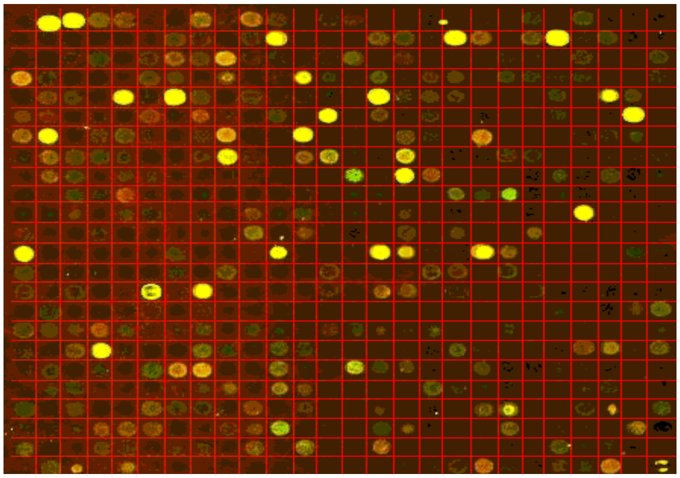


Figure 12. Gridding of noisy microarray image after refinement process

5. Conclusion

In this paper, we have presented a new spatial topology technique to automatically grid skewed noisy microarray images. The results of our experiment on skewed microarray images on Stanford databases and UNC are encouraging. The skew correction algorithm depends on determination of coordinate addresses of just two positions of the image. The noise removal is done through adaptive thresholding which makes processes effective. Finally the gridding is performed based on spatial topology of spots. To summarize the three stages of the proposed method when executed in sequence is effective and computationally simple.

References

[1] M. B. Eisen, ScanAlyze Nov. 1999 [Online]. Available: http://rana.lbl. gov/EisenSoftware.htm

[2] J. Buhler, T. Ideker, and D. Haynor, Dapple: Improved techniques for finding spots on DNA microarrays UW CSE Tech. Rep. UWTR 2000-08-05, Aug. 2000, pp. 1–12.

[3] Biodiscovery, Inc., ImaGene 2005 [Online]. Available: http://www. biodiscovery.com/imagene.asp

[4] P. Hegde *et al.*, "A concise guide to cdna microarray analysis," *Biotechniques*, vol. 29, no. 3, pp. 548–556, Sep. 2000.

[5] A. N. Jain, T. Tokuyasu, A. Snijders, R. Segraves, D. Albertso, and D. Pinkel, "Fully automatic quantification of microarray image data," *Genome Res.*, vol. 12, pp. 325–332, 2003.

[6] A. W. Liew, H. Yan, and M. Yang, "Robust adaptive spot Segmentation of DNA microarray images," *Pattern Recognit.*, vol. 36, pp. 1251–1254, 2003.

[7] K. Hartelius and J. M. Carstensen, "Bayesian grid matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 2, pp. 162–173, Feb. 2003.

[8] M. Katzer, F. Kummert, and G. Sagerer, "A Markov random field model of microarray gridding," presented at the ACM Symp. Applied Computing 2003.

[9] N. Deng and H. Duan, "The automatic gridding algorithm based on Projection for microarray image," in *Proc. 2004 Int. Conf. Intell. Mechatronics Automation*, Chengdu, China, 2004, pp. 254–257.

[10] A. W. Liew, H. Yan, and M. Yang, "Robust adaptive spot Segmentation of DNA microarray images," *Pattern Recognit.*, vol. 36, pp. 1251–1254, 2003.

[11] L. Rueda and V.Vidyadharan, "A hill-climbing approach for automatic gridding of cdna microarray images," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 3, no. 1, pp. 72–83, Jan.-Mar. 2006.

[12] G. Antoniol and M. Ceccarelli, "A markov random field approach to microarray image gridding," in *Proc. IEEE Int. Conf. Pattern Recognit.*, Cambridge, U.K., 2004, pp. 550–553.

[13] H. Y. Jung and H. G. Cho, "An automatic block and spot indexing with k-nearest neighbors graph for microarray image analysis," *Bioinformatics*, vol. 18, no. 1, pp. 141–151, Oct. 2002.

[14] P. Bajcsy, "Gridline: Automatic grid alignment in DNA microarray scans," *IEEE Trans. Image Process.*, vol. 13, no. 1, pp. 15–25, Jan. 2004.

[15] M. Steinfath, W. Wruck, H. Seidel, H. Lehrach, U. Radelof, and J. O'Brien, "Automated image analysis for array hybridization experiments," *Bioinformatics*, vol. 17, no. 7, pp. 634–641, Jul. 1, 2001.

[16] N. Brandle, H. Bischof, and H. Lapp, "Robust DNA microarray image analysis," *Mach. Vis. Appl.*, vol. 15, pp. 11–28, 2003.

[17] J. Ho, W. L. Hwang, H. H. S. Lu, and D. T. Lee, "Gridding spot centers of smoothly distorted microarray images," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 342–353, Feb. 2006.

[18] G. Antoniol and M. Ceccarelli, "Microarray image gridding with stochasticsearch based approaches," *Image Vison Comput.*, vol. 25, no. 2, pp. 155–163, Feb. 2007.

AUTHORS PROFILE

Manjunath S.S has received B.E degree in 2000 from Mysore University, Mysore and M.Tech degree in 2005 from VTU University, Belgaum, Karnataka, India. Currently he is working as a Assistant Professor at Dayananda Sagar College of Engineering, Karnataka, India and His experience in teaching started from the year 2000. Currently his pursuing PhD in mysore university. His areas of interests include microarray image processing, medical image segmentation and clustering algorithms.

Dr. Lalitha Rangarajan has received Master degrees in Mathematics from Madras University, India and from the Department of Industrial Engineering, Purdue University. She completed Ph.D in Computer Science from University of Mysore, India. She has been teaching courses in Mathematics, Operations Research and Computer Science, for Master degree students for more than 25 years. She is presently a Reader at Department of Computer Science, University of Mysore, India. Her current research interests are Image Retrieval, Feature Reduction and Bio Informatics. She has more than 40 publications in reputed journals and conferences.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

(IJCSIS) International Journal of Computer Science and Information Security,
Vol. 8, No. 4, July 2010

LDCP+: An Optimal Algorithm for Static Task Scheduling in Grid Systems

Negin Rzavi
Islamic Azad University,
Science and Research Branch,
Tehran, Iran
n.razavi@srbiau.ac.ir

Safieh Siadat
Islamic Azad University,
Science and Research Branch,
Tehran, Iran
s.siadat@srbiau.ac.ir

Amir Masoud Rahmani
Islamic Azad University,
Science and Research Branch,
Tehran, Iran
rahmani@srbiau.ac.ir

***Abstract*— after a computational job is designed and realized as a set of tasks, an optimal assignment of these tasks to the processing elements in a given architecture needs to be determined. In grid system with the existence of heterogeneous processing elements and data transferring time between them, determining an assignment of tasks to processing elements in order to optimize the performance and efficiency is so important. In this paper a heuristic algorithm named LDCP+ is presented, which has optimized the Longest Dynamic Critical Path algorithm (LDCP) presented by Mohammad L. Daoud and Nawwaf Kharma in 2007. This algorithm is a list-based algorithm in the way it assigns each task a priority for its execution. Using task duplication, using idle processing element's time and also optimizing priority assignment method which is used in LDCP algorithm, are the basic specifications of LDCP+, since LDCP algorithm is executable with the assumption that computation cost of tasks are monotonic, our algorithm which is presented in this paper has made the scheduling algorithm free from this restriction and in the case of non-monotonic computation costs, LDCP+ has the minimum total finish time in the comparison of other scheduling algorithms such as HEFT and CPOP.**

Keywords- Grid; Static task scheduling; Longest Dynamic Critical Path.

I. INTRODUCTION

A Grid system is a group of connected computers that has the ability of executing parallel programs via a high speed interconnection. The efficiency of program parallelism in Grid systems depends on methods used in task scheduling on available processing elements. Inner connection of processing elements in Grid causes an overhead when two tasks assigned to different processing elements of distinct computers, transfer data. In fact, task scheduling in distributed heterogeneous systems are more complex in which each task can have different execution time on different processing elements, so scheduling algorithms for a Grid system should consider the execution time of each task on different processing elements and even one incorrect decision can restrict the system performance to the slowest processing element [2].

There are two kinds of scheduling algorithms: static scheduling algorithms and dynamic scheduling algorithms. In static scheduling algorithms all information needed for scheduling such as the structure of the parallel application, the execution time of individual tasks and the communication costs between tasks must be known, in contrast, these information are unknown in dynamic task scheduling algorithms.

Among different types of scheduling algorithms, HEFT is a scheduling algorithm for heterogeneous distributed computing systems which is consists of two phases: first, cost computing for each task and task selection, second, processor selection. In the task selection phase the algorithm sets the computation costs of tasks to their mean values and this may limit the ability of scheduling algorithm to precisely compute the priorities of tasks. The CPOP algorithm is same as HEFT in the two phases but with different strategies in assigning priorities to tasks and processor selection. These two algorithms have been mentioned as optimal algorithms in the parameter of total finish time.

In this paper we present a heuristic list-based algorithm called LDCP+ (optimized of Longest Dynamic Critical Path algorithm) for static task scheduling in Grid systems with limited number of processors and we compare our scheduling results with other algorithms such as CPOP, HEFT and LDCP for performance evaluation.

II. RELATED WORKS

Static task scheduling for Grid systems, in general is known to be NP-Complete problem [4, 7, 9] and most of these algorithms are heuristic [1, 2, 3, 4, 7]. One of the most important classes of heuristic algorithms is list-based algorithms [6], in such algorithms each task is assigned with a priority and three steps of task selection, processor selection and status update are repeated until all tasks are scheduled. In the task selection phase the unscheduled task with the highest priority is selected. In the processor selection phase, the selected task is assigned to the processor that minimizes a predefined cost criterion that can be minimizing the schedule length. At last in status update phase, the status of the system is updated. Examples of list-based algorithms are: Heterogeneous Earliest Finish Time (HEFT) [9], Critical Path on a Processor (CPOC) [9], Critical Path on a Cluster (CPOC) [5], Dynamic

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Level Scheduling (DLS) [8], Modified Critical Path (MCP) [10], Mapping Heuristic (MH) [3], Dynamic Critical Path (DCP), and Longest Dynamic Critical Path (LDCP) [2].

III. Problem Definition

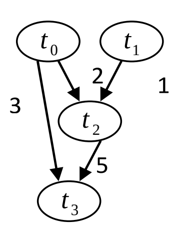
In static task scheduling in Grid system, the execution precedence between tasks is represented by a Directed Acyclic Graph (DAG), each DAG is shown by tupple (T, E) where T is a set of n tasks and E is a set of e edges. Each $t_i \in T$ represents a task and each $e_{i,j} = (t_i, t_j) \in E$ represents the execution precedence between the two tasks which are connected with the edge $e_{i,j}$.

If $(t_i, t_j) \in E$ then the execution of task $t_i \in T$ cannot be started before task finishes its execution. For the edge (t_i, t_j), the source task t_i is parent of the sink task t_j, while t_j is a child of t_i. A task with no parents is called an entry task and a task with no children is called an exit task. Associated with each edge (t_i, t_j) is a value $d_{i,j}$ that represents the amount of data to be transmitted from task to task t_j, and in some cases it also represents the minimum time that a task needs to wait for starting after task t_i finishes its execution.

A Grid system is represented by a set *P* of *m* processors, a set *T* of *n* tasks and $n \times m$ computation cost matrix ($W_{n \times m}$). Each element $w_{i,k} \in W, 1 \le i < n, 1 \le k \le m$ represents the execution time (computation cost) of task t_i on processor P_k. We have the same assumption as LDCP that all processors are fully connected and communications between processors occur via independent communication units [2], so, we can have task execution and data transferring in parallel. Also the data transfer rate between any two processors on the network is assumed to be fixed and constant as same as LDCP. The communication cost between two processors is represented by $n \times n$ matrix ($D_{n \times n}$). $d_{i,j} \in D$ is zero if two tasks t_i and t_j of and are scheduled on the same processor and it is equal to communication cost (non zero) in the other case. A task can start its execution on a processor only when all data from its parent become available to that processor. The goal of our algorithm is to assign tasks in processors in a way that minimizes the total finish time, or the schedule length.

A. Definition 1

Schedule Length: The maximum execution time of the processors or the finish time of the final task after task scheduling is called scheduled length. There is a DAG and a computation cost matrix with two processors as shown in Fig.1. The schedule length is computed in Fig.2. and it is equal to 23.



(a)

Task	p_0	p_1
t_0	6	5
t_1	6	4
t_2	9	8
t_3	7	3

(b)

Figure 1. An example of (a) DAG (b) computation matrix

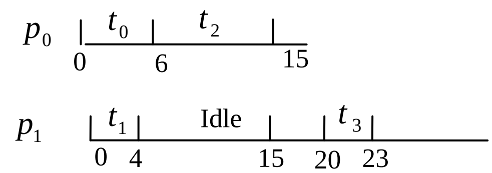
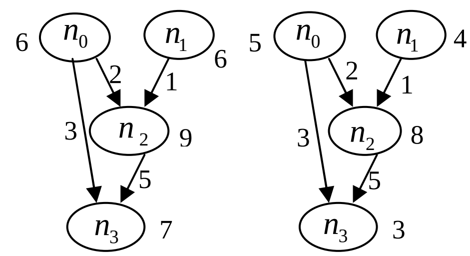


Figure 2. Schedule length of the presented DAG in Fig. 1. on two processors



(a) (b)

Figure 3. Tasks computation time on each processor that will be acquired from cost matrix in Fig. 1.

Assigning task priorities in Grid system the efficiency of list-based scheduling algorithms depends on the methods which assign priorities to tasks.

In our suggested algorithm LDCP+, if selecting a task in one step of scheduling causes the minimum schedule length we assign a high priority to that task. There are some basic definitions which are used in LDCP algorithm and because LDCP+ is the result of optimization of LDCP, we decided to represent this basic knowledge too.

B. Definition 2

Critical Path: For a given DAG, the Critical Path (CP) is defined as the path from an entry task to an exit task for which the sum of the computation costs of tasks and the communication costs of edges is maximal.

IV. LDCP: Longest Dynamic Critical Path

A. Definition 3

Longest Dynamic Critical Path: Given a DAG with n tasks and e edges and a Grid system with m processors, DCP during a particular scheduling step is a path of tasks and edges from an entry task to an exit task.

LDCP is the largest DCP, considering that communication costs between tasks scheduled on the same processors are assumed zero, and the execution constraints are preserved. Fig.3. represents two dynamic critical paths. First path in Fig.3.a. is composed of tasks t_0, t_2 and t_3 which is scheduled on processor p_0 and has the length of 29. The second DCP in Fig.3.b. is composed of tasks t_0, t_2 and t_3 which is scheduled on processor p_1 and has the length of 23, so at the first step of scheduling, LDCP is composed of tasks t_0, t_2 and t_3 and with the schedule length of 29.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

V. LDCP+: The Proposed Algorithm

In the algorithm of LDCP+, each scheduling iteration includes three phases below:

1. Task selection

2. Processor selection phase

3. Status update phase

These 3 phases will be accomplished for each task until last input task is selected for scheduling.

A. Task Selection Phase

LDCP+ selects a set of tasks that play main role in determining schedule length.

In first step of this phase, DAG of each processor is required for scheduling.

1) Definition 4

Directed Graph: With the assumption of having a DAG including *n* tasks, e edges and a Grid system with *m* processors ($p_0, p_1, ..., p_m$), $DAGP_k$ is the directed acyclic graph that corresponds to processor p_k . The computation cost of each task in the processor p_k , is represented by a number on the related node of the $DAGP_k$.

$DAGP_0$ is shown in Fig.3.a. and $DAGP_1$ is shown in Fig.3.b. These figures are related with the DAG and the Grid system which is represented in Fig.1. Trough the course of this paper, t_i is used to refer to the *i'th* task in directed acyclic graph and the node n_i identifies task t_i in $DAGP_k$. The number associated with this node represents the computation cost of task ti on processor pk. In each $DAGP_k$, all nodes are assigned with a number named UpwardRank (URank). URanks are used to determine tasks priorities in $DAGP_k$.

2) Definition 5

URank: UpwardRank of *i'th* node (n_i) in $DAGP_k$ is defined recursively as

$$URank_k(n_i) = w_{i,k} + \max_{n_l \in succ_k(n_i)}\{c_k(n_i, n_l) + URank(n_l)\} \quad (1)$$

where $succ_k(n_i)$ is a set of immediate successors of node n_i , $c_k(n_i, n_l)$ is the communication cost between nodes n_i and n_l in $DAGP_k$, and $w_{i,k}$ is the computation cost of t_i on processor p_k .

3) Definition 6

URankSet: Each element of URankSet is defined as

$$Max\{\sum_{k=0}^{m-1} URank_k(n_i)\} \quad (2)$$

where $URank_k(n_i)$ is $URank(n_i)$ in $DAGP_k$.

4) Definition 7

KeyNode: KeyNode is the node that has the maximum URank in URankSet. Corresponded task to this node is used as selected task for scheduling algorithm.

5) Definition 8

KeyNodeSet: This set includes KeyNodes that are selected for scheduling and in the first scheduling iteration it can include several tasks, but in other iterations it has only one task for scheduling and in the first scheduling iteration it can include several tasks, but in other iterations it has only one task.

6) Definition 9.

Least Execution Time (LET): Least Execution Time is defined as

$$\min\{processTime(p_k) + w_{i,k} + d_{i,j}\} \quad (3)$$

where $processTime(p_k)$ is the time that find scheduled task on processor p_k finishes its execution, $w_{i,k}$ is the computation time of task corresponded to *i* on processor *k*, and $d_{i,j}$ is the communication time between t_i and t_j . If both t_i and t_j are scheduled on processor p_k , then communication cost between them will be assumed zero. After computing URankSet, the destined task for scheduling algorithm is the task corresponding to existing KeyNode in URankSet. In the first iteration to obtain minimum execution time on available processing elements, if the number of entry tasks is equal or less than processors number, all entry tasks will be consider as KeyNode, in other case, as same as the number of processors, tasks with maximum URanks will be selected as KeyNodes and place in KeyNodeSet. In the next iterations, KeyNodeSet merely includes one KeyNode (a set with one member).

B. Processor Selection Phase

In this phase, selected task will be assigned to a processor in the way to gain the minimum schedule length in each iteration of scheduling. Therefore, in LDCP+, these stages will be passed: As mentioned above, in the first scheduling iteration, KeyNodeSet can have more than one KeyNode. For the purpose of optimizing LDCP algorithm, LDCP+ computes distinct permutation of tasks, which their corresponding KeyNodes are available in KeyNodeSet, on different processors and the permutation with the minimum average execution time on processors will be the first assignment of tasks to processors. This average execution time can be achieved from

$$\min\left\{\frac{\sum_{k=0}^{m-1} w_{i,k}}{m}\right\} \quad (4)$$

Where *i* is the number of tasks corresponding to their KeyNodes, $w_{i,k} \in w$ and *m* is the number of processors. In the next iterations, the only available KeyNode in KeyNodeSet is selected to be scheduled.

1) Definition 10

Idle Space: In a processor when there is a gap between the start time of a task and the end time of the previous task, that interval time is called idle space.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

2) Definition 11

Replacement Ability: One task can be placed in an idle space when parents of that task have been terminated before the start time of the task. If any of its parents have been scheduled on a different processor, the required time for transferring data between processors should be mentioned.

If there is a processor with the idle space and selected task has the ability of locating in that space (replacement ability), that processor will be selected. At the end of this phase, LDCP+ algorithm uses duplication process to decrease the schedule length. With this definition, after selecting the processor if the selected task has a parent scheduled on a different processor and the selected processor has an idle space before the start time of the selected task, then duplication process in the idle space will be used (regarding to the replacement ability).

3) Definition 12

Duplication Process: Duplication process is repeating the execution of one task on other processors.

C. Status Update Phase

After selecting the task and assigning it to a processor, appropriate URank with the selected task will be deleted from URankSet. Finish process time of the selected processor will be updated after the task has been assigned to the processor. Selected task will be deleted from the list of unscheduled tasks. LDCP+ algorithm is proposed in Fig.4.

VI. CASE STUDY

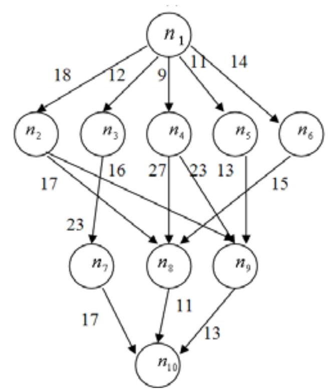
In this section, execution results of CPOP, HEFT and LDCP+ algorithms are compared in the case of having non monotonic computation cost matrix. A Grid system compose of three single-processor computers (*m*=3), fifteen tasks (*n*=15), a non monotonic computation cost matrix and a DAG with communication costs assigned to graph edges are shown in Fig.5. which also presents scheduling results of the mentioned DAG, executed by HEFT, CPOP and LDCP+ algorithms. Execution results of LDCP and LDCP+ are compared according to monotonic computation cost matrix. A Grid system with the parameters *m*=2 and *n*=10, a monotonic computation cost matrix and a DAG with communication costs assigned to graph edges are shown in Fig.6. Fig.6 also shows scheduling results of the mentioned DAG presented in Fig6.b, executed by LDCP and LDCP+ scheduling algorithms.

Establish $DAGP_k$ for all processors in the system where $0 \le k \le m-1$
Calculate URanks for all $DAGP_k$
Compute the URankSet
While there are unscheduled tasks in task list **do**
  Find the KeyNode(s) in the URankSet
  Put the KeyNode(s) in KeyNodeSet
  If (it's the first step of scheduling) **then**
    Choose the processors which have the minimum permutation;
  Else
    If (there is any processor with idle time and the task have the replacement ability) **then**
      Selected the processor;
    Else
      Compute the finish time of the selected task on every system;
      Find and select the processor that minimizes the finish time of the Selected task;
    End if
      Duplicate the parent(s) of the selected task if needed;
  End if
  Assign the selected task to the selected processor;
  Update the selected processor time;
  Update the URANK set;
  Update unscheduled task list;
End while

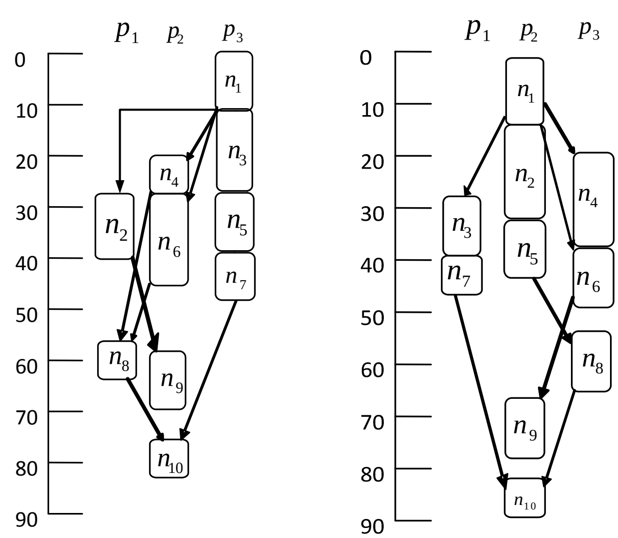
Figure 4. LDCP+ algorithm

Task	p_1	p_2	p_3
1	14	16	9
2	13	19	18
3	11	13	19
4	13	8	17
5	12	13	10
6	13	16	9
7	7	15	11
8	5	11	14
9	18	12	20
10	21	7	16

(a)



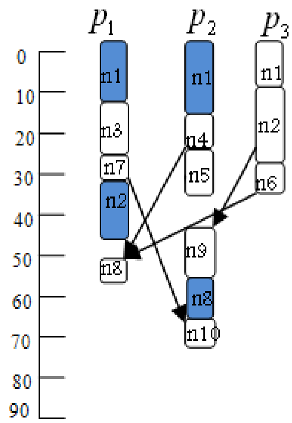
(b)



(c) (d)

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

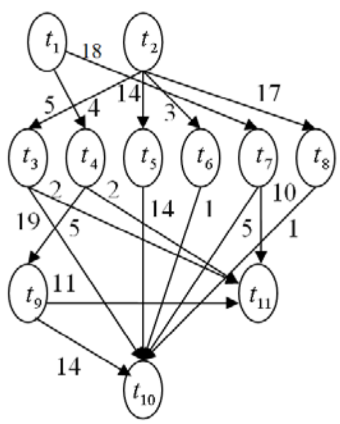
step	Selected task	Selected processor
1	t_2	p_0
2	t_1	p_1
3	t_4	p_1
4	t_9	p_0
5	t_5	p_0
6	t_3	p_0
7	t_7	p_1
8	t_6	p_1
9	t_8	p_0
10	t_{11}	p_0
11	t_{10}	p_1



(e) (f)

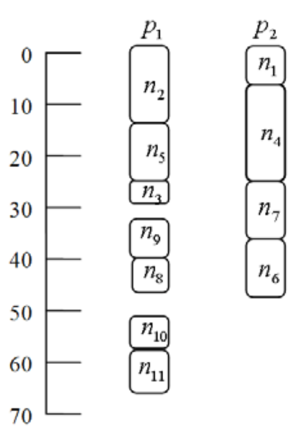
Figure 5. Scheduling results for HEFT, CPOP, LDCP+ algorithms. (a) A graph with 10 tasks. (b) Graph cost matrix. (c) HEFT Scheduling algorithm with schedule length of 80. (d) CPOP algorithm with schedule length of 89. (e) LDCP+ algorithms with schedule length of 68. (f) Tasks execution sequence in LDCP+ algorithm. **Duplicated tasks:**

task	p_0	p_1
t_1	4	6
t_2	15	22.5
t_3	4	6
t_4	13	19.5
t_5	10	15
t_6	7	10.5
t_7	8	12
t_8	4	6
t_9	12	18
t_{10}	6	9
t_{11}	9	13.5



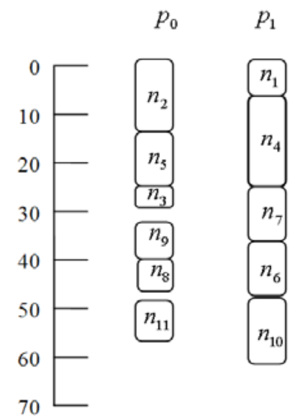
(a) (b)

step	Selected task	Selected processor
1	t_2	p_0
2	t_1	p_1
3	t_4	p_1
4	t_9	p_0
5	t_5	p_0
6	t_3	p_0
7	t_7	p_1
8	t_6	p_1
9	t_8	p_0
10	t_{10}	p_0
11	t_{11}	p_0



(c) (d)

step	Selecte d task	Selected processo r
1	n_1	p_1
2	n_4	p_2
3	n_3	p_1
4	n_2	p_3
5	n_5	p_2
6	n_6	p_3
7	n_9	p_2
8	n_7	p_1
9	n_8	p_1
10	n_{10}	p_2



(e) (f)

Figure 6. Scheduling results for LDCP and LDCP+ algorithms. (a) A graph with 11 tasks. (b) Graph cost matrix (c) tasks execution sequence in LDCP algorithm (d) LDCP algorithm with schedule length of 64 (e) tasks execution sequence in LDCP+ algorithm. (f) LDCP+ algorithm with schedule length of 61.5

VII. Conclusion and Future Work

In Grid systems, task scheduling is an important problem in the domain of optimizing heterogeneous distributed systems. In this paper a new heuristic scheduling algorithm, named LDCP+, is proposed. This algorithm has optimized LDCP algorithm that better result are attained for schedule length by improving these three phases: task selection phase, processor selection phase and status update phase. LDCP+ can schedule tasks in Grid systems in both case of having monotonic and non monotonic cost matrix. Using duplication process for optimizing priority assigns to tasks and also using idle spaces of processors will result in having better schedule length rather than other scheduling algorithms. In real time environment, the assignment of resources such as processors in a specific time is so important. More works can be done to improve algorithms with less computation cost for such environments.

References

[1] S. Bansal, P. Kumar, and K. Singh. An improved duplication strategy for scheduling precedence constrained graphs in multiprocessor systems. In IEEE Transactions on Parallel and Distributed Systems 14(6), pages 533-544, 2003.

[2] M. I. Daoud and N. N. Kharma. A high performance algorithm for static task scheduling in heterogeneous distributed computing systems. In Journal of Parallel and Distributed Computing 68(4), pages 399-409, 2008.

[3] H. El-Rewini and T. G. Lewis. Scheduling parallel program tasks onto arbitrary target machines. Journal of Parallel and Distributed Computing 9(2), pages 138-153, 1990.

[4] E. Ilavarasan, P. Thambidurai, and R. Mahilmannan. Performance effective task scheduling algorithm for heterogeneous computing system. 4th International Symposium on Parallel and Distributed Computing, 0:28-38, 2005.

[5] J. Kim, J. Rho, J.-O. Lee, and M.-C. Ko. Cpoc: Effective static task scheduling for grid computing. In Proceeding of the 2005 International

[6] Conference on High Performance Computing and Communcations, pages 477-486, 2005.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[7] Y.-K. Kwok and I. Ahmad. Static scheduling algorithms for allocating directed task graphs to multiprocessors. ACM Comput. Surv. 31(4), pages 406-471, 1999.

[8] Y. kwong Kwok, I. Ahmad, and I. Ahmad. Dynamic critical-path scheduling: An effective technique for allocating task graphs to multiprocessors. IEEE Transactions on Parallel and Distributed Systems 7(5), pages 506-521, 1996.

[9] G. C. Sih and E. A. Lee. A compile-time scheduling heuristic for interconnection constrained heterogeneous processor architectures. IEEE Transaction on Parallel and Distributed Systems 4(2), pages 175-187, 1993.

[10] H. Topcuoglu, S. Hariri, and W. Min-You. Performance-effective and low complexity task scheduling forheterogeneous computing. *IEEE Transaction on Parallel and Distributed Systems 13(3)*, pages 260-274, 2002.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Density Distribution and Sector Mean with Zero-cal and Highest-sal Components in Walsh Transform Sectors as Feature Vectors for Image Retrieval

H.B.Kekre
Sr. Professor
MPSTME, SVKM's NMIMS (Deemed-to be-University)
Vile Parle West, Mumbai -56,INDIA
hbkekre@yahoo.com

Dhirendra Mishra
Assistant Professor & PhD Research Scholar
MPSTME, SVKM's NMIMS (Deemed-to be-University)
Vile Parle West, Mumbai -56,INDIA
dhirendra.mishra@gmail.com

Abstract- **We have introduced a novel idea of considering complex Walsh transform for sectorization of transformed components. In this process the first coefficient of zero-cal and the last coefficient highest-sal are not used. In this paper we have proposed two different approaches along with the extra components of zero-cal and highest-sal for feature vector generation namely sector density and sector mean. Two similarity measures such as sum of absolute difference and Euclidean distance are used and results are compared. The cross over point performance of overall average of precision and recall for both approaches on different sector sizes are compared. The density distribution of real (cal) and imaginary (sal) values and sector mean of Walsh sectors in all three color planes are considered to design the feature vector. The algorithm proposed here is worked over database of 1055 images spread over 12 different classes. Overall Average precision and recall is calculated for the performance evaluation and comparison of 4, 8, 12 & 16 Walsh sectors. The use of sum of absolute difference as similarity measure always gives lesser computational complexity and density distribution approach with sum of absolute difference as similarity measure of feature vector has the best retrieval performance.**

Keywords-CBIR, Walsh Transform, Euclidian Distance, Absolute Difference, Precision, Recall

I. INTRODUCTION

With the huge growth of digital information the need of its management requires need of storage and utilization in efficient manner. This has lead to approach like content based image search and retrieval to be used. The earliest use of the term content-based image retrieval in the literature seems to have been by Kato [1], to describe his experiments into automatic retrieval of images from a database by color, texture and shape features. The term has since been widely used to describe the process of retrieving desired images from a large collection on the basis of features (such as colors, texture and shape) that can be automatically extracted from the images themselves. The typical CBIR system [1-6] performs two major tasks. The first one is feature extraction (FE), where a set of features, called image signature or feature vector, is generated to accurately represent the content of each image in the database. A feature vector is much smaller in size than the original image, typically of the order of hundreds of elements (rather than millions). The second task is similarity measurement (SM), where a distance between the query image and each image in the database using their signatures is computed so that the top closest images can be retrieved.[7-9]. There are various approaches which have been experimented to generate the efficient algorithm for CBIR like FFT sectors [4-6], Transforms [15][17], Vector quantization[12], bit truncation coding [13][14]. In this paper we have introduced a novel concept of complex Walsh transform and its sectorization for feature extraction (FE).Two different similarity measures namely sum of absolute difference and Euclidean distance are considered. The performances of these approaches are compared.

II. WALSH TRANSFORM

Walsh transform [17] matrix is defined as a set of N rows, denoted Wj, for j = 0, 1, ... , N - 1, which have the following properties:

- Wj takes on the values +1 and -1.
- Wj[0] = 1 for all j.
- Wj x W^T_k=0, for j ≠ k and Wj x Wk^T =N, for j=k.
- Wj has exactly j zero crossings, for j = 0, 1, ., N-1.
- Each row Wj is either even or odd with respect to its midpoint.

Walsh transform matrix is generated using a Hadamard matrix of order N. The Walsh transform matrix row is the row of the Hadamard matrix specified by the Walsh code index, which must be an integer in the range [0, ..., N - 1]. For the Walsh code index equal to an integer j, the respective Hadamard output code has exactly j zero crossings, for j = 0, 1, ... , N - 1.

Kekre's Algorithm to generate Walsh Transform from Hadamard matrix [17] is illustrated for N=16.However the algorithm is general and can be used for any $N = 2^k$ where k is an integer.

Step 1:

Arrange the 'N' numbers in a row and then split the row at 'N/2', the other part is written below the upper row but in reverse order as follows:

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

15 14 13 12 11 10 9 8

Step 2:

We get two rows, each of this row is again split in 'N/4' and other part is written in reverse order below the upper rows as shown below.

0 1 2 3

15 14 13 12

7 6 5 4

8 9 10 11

This step is repeated until we get a single column which gives the ordering of the Hadamard rows according to sequency as given below:

0 ,15, 7, 8, 3,12,4,11,1,14,6,9,2,13,5,10

Step 3:

According to this sequence the Hadamard rows are arranged to get Walsh transform matrix. Now a product of Walsh matrix and the image matrix is calculated. This matrix contains Walsh transform of all the columns of the given image.

Since Walsh matrix has the entries either +1 or -1 there is no multiplication involved in computing this matrix. Since only additions are involved computational complexity is very low.

III. FEATURE VECTOR GENERATION

The proposed algorithm makes novel use of Walsh transform to design the sectors to generate the feature vectors for the purpose of search and retrieval of database images. The complex Walsh transform is conceived by multiplying all sal functions by $j = \sqrt{-1}$ and combining them with real cal functions of the same sequency. Thus it is possible to calculate the angle by taking $\tan^{-1}$ of sal/cal. However the values of tan are periodic with the period of π radians hence it can resolve these values in only two sectors. To get the angle in the range of 0-360 degrees we divide these points in four sectors as explained below. These four sectors are further divided into 8, 12 and 16 sectors. We have proposed two different approaches for feature vector generation namely density and mean value of all the vectors in each sector with sum of absolute difference and Euclidean distance [7-9] [11-14] as similarity measures. Performance of both these approaches are compared using both similarity measures. Each Walsh sector is represented by single percentage value of sal-cal distribution in particular sector w.r.t. all sectors for feature vector generation using formula shown in (1) and two extra components of zero-cal and highest-sal are added to the feature vector.

$$\frac{\text{(Total number of sal in particular sector)}}{\text{(Total number of sal in all sectors)}} \times 100 \qquad (1)$$

Thus for 4, 8, 12 & 16 Walsh sectors 4, 8, 12 and 16 feature components along with extra components of zero-cal and highest-sal for each color planes i.e. R, G and B are generated. Thus all feature vectors are of dimension 18, 30, 42 and 54 components. In the second approach mean of all sal and cal of each sector and with extra components of zero-cal and highest-sal is calculated and equation (2) is used for feature component representing each sector for all three color planes i.e. R. G and B with single value thus forming the feature vector of dimension 18, 30, 42 and 54 for 4, 8, 12 and 16 complex Walsh transform sectors.

A. Four Walsh Transform Sectors:

To get the angle in the range of 0-360 degrees, the steps as given in Table 1 are followed to separate these points into four quadrants of the complex plane. The Walsh transform of the color image is calculated in all three R, G and B planes. The complex rows representing sal components of the image and the real rows representing cal components are checked for positive and negative signs. The sal and cal Walsh values are assigned to each quadrant. as follows:

TABLE I. FOUR WALSH SECTOR FORMATION

Sign of Sal	**Sign of Cal**	**Quadrant Assigned**
+	+	I (0 – 90 Degrees)
+	-	II (90 – 180 Degrees)
-	-	III(180- 270 Degrees)
-	+	IV(270–360 Degrees)

The equation (1) is used to generate individual components to generate the feature vector of dimension 12 considering three R, G and B Planes in the sal and cal density distribution approach. However, it is observed that the density variation in 4 quadrants is very small for all the images. Thus the feature vectors have poor discretionary power and hence higher number of sectors such as 8, 12 and 16 were tried. In the case of second approach of feature vector generation i.e. individual sector mean has better discretionary power in all sectors and equation (2) is used to generate the individual sector components.

$$\sqrt{(\text{mean of sal vector})^2 + (\text{mean of cal vector})^2} \qquad (2)$$

Sum of absolute difference measure is used to check the closeness of the query image from the database image and

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

precision and recall are calculated to measure the overall performance of the algorithm.

B. *Eight Walsh Transform Sectors:*

Each quadrants formed in the previous obtained 4 sectors are individually divided into 2 sectors each considering the angle of 45 degree. In total we form 8 sectors for R,G and B planes separately as shown in the Table 2. The percentage density distribution of sal and cal in all 8 sectors are determined using equation (1) to generate the feature vector.

TABLE 2. EIGHT WALSH SECTOR FORMATION

Quadrant of 4 Walsh sectors	**Condition**	**New sectors Formed**
I (0 – 90^0)	Cal >= Sal	I (0-45 Degrees)
	Sal > Cal	II (45-90 Degrees)
II (90 – 180^0)	|Sal | > |Cal|	III(90-135 Degrees)
	|Call >= |Sal|	IV(135-180 Degrees)
III (180- 270^0)	|Call >= |Sal|	V (180-225 Degrees)
	|Sal| > |Cal|	VI (225-270 Degrees)
IV (270 – 360^0)	|Sal| > |Cal|	VII (270-315 Degrees)
	|Call >= |Sal|	VIII (315-360 Degrees)

C. *Twelve Walsh Transform Sectors:*

Each quadrants formed in the previous section of 4 sectors are individually divided into 3 sectors each considering the angle of 30 degree. In total we form 12 sectors for R,G and B planes separately as shown in the Table 3. The percentage density distribution of sal and cal in all 12 sectors are determined using equation (1) to generate the feature vector

TABLE 3. TWELVE WALSH SECTOR FORMATION

4 Quadrants	**Condition**	**New sectors**
I (0 – 90^0)	Cal >= √3 * Sal	I (0-30 ^{0})
	1/√3 cal <=sal<= √3 cal	II (30-60 ^{0})
	Otherwise	III (60-90 ^{0})
II (90 – 180^0)	Cal >= √3 * Sal	IV (90-120 ^{0})
	1/√3 |cal| <=|sal|<= √3 |cal|	V (120-150 ^{0})
	Otherwise	VI (150-180 ^{0})
III (180- 270^0)	|Cal|>= √3 * |Sal|	VII (180-210 ^{0})
	1/√3 cal <=|sal|<= √3 |cal|	VIII(210- 240 ^{0})
	Otherwise	IX (240-270 ^{0})
IV (270 – 360^0)	|Cal|>= √3 * |Sal|	X (270-300 ^{0})
	1/√3 |cal| <=|sal|<= √3 |cal|	XI (300-330 ^{0})
	Otherwise	XII (330-360 ^{0})

IV. RESULTS AND DISCUSSION

The sample Images of the database of 1055 images of 12 different classes such as Flower, Sunset, Barbie, Tribal, Puppy, Cartoon, Elephant, Dinosaur, Bus, Parrots, Scenery,Beach is shown in the Figure 1.

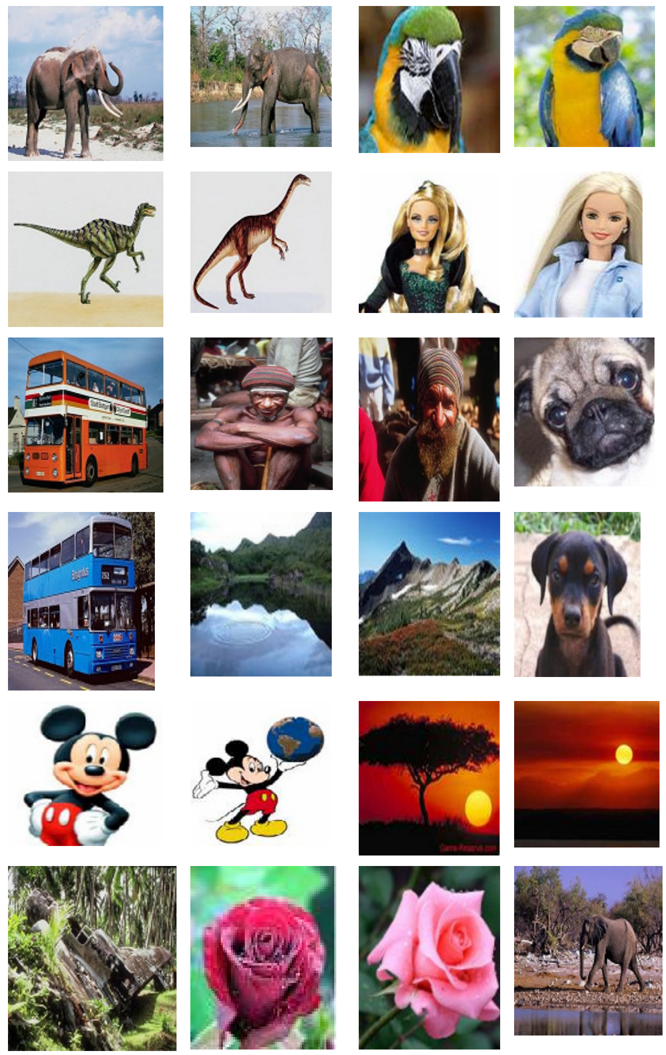


Figure 1. Sample Image Database

Figure 2. Query Image

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

The flower class image is taken as sample query image as shown in the Figure 2 for both approaches of sal cal density distribution and individual sector mean. The first 21 images retrieved in the case of sector mean in 12 Walsh sector used for feature vectors and sum of absolute difference as similarity measure is shown in the Figure 3. It is seen that only 4 images of irrelevant class are retrieved among first 21 images and rest are of query image class i.e. flower. Whereas in the case of sal cal density in 12 Walsh Sectors with sum of absolute difference as similarity measures there is only 1 image of irrelevant class and 20 images of the query class i.e. flower are retrieved as shown in the Figure4.

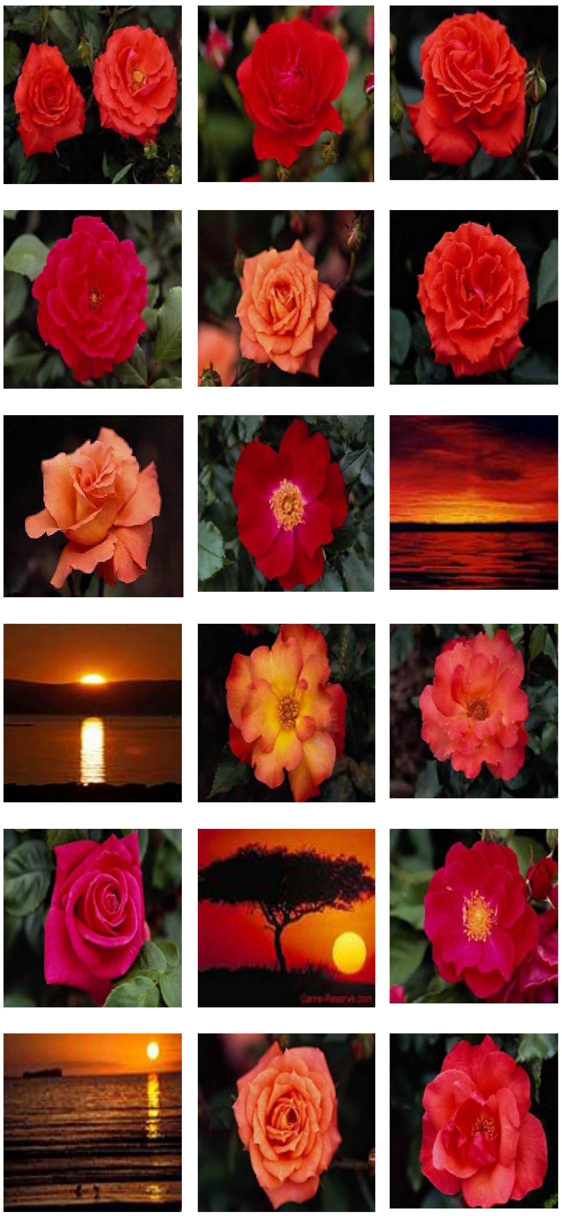
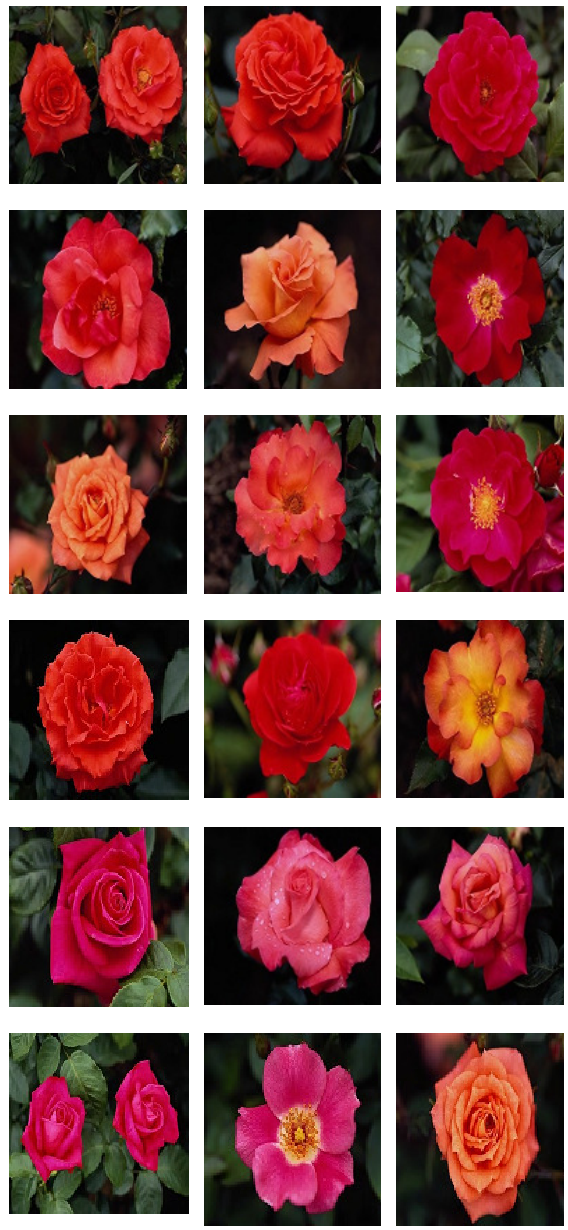


Figure 3: First 21 Retrieved Images based on individual sector mean of 12 Walsh Sectors with Absolute Difference as similarity measures for the query image shown in the Figure 2.



http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Figure 4: First 21 Retrieved Images based on sal cal density distribution in 12 Walsh Sectors with Absolute Difference as similarity measures for the query image shown in the Figure 2.

Once the feature vector is generated for all images in the database a feature database is created. A query image of each class is produced to search the database. The image with exact match gives minimum absolute difference. To check the effectiveness of the work and its performance with respect to retrieval of the images we have calculated the precision and recall as given in Equations (3) & (4) below:

$$\text{Precision} = \frac{\text{Number of relevant images retrieved}}{\text{Total Number of images retrieved}} \qquad (3)$$

$$\text{Recall} = \frac{\text{Number of relevant images retrieved}}{\text{Total number of relevant images in database}} \qquad (4)$$

The Figure5 - Figure8 shows the Overall Average Precision and Recall performance of sal cal density in 4, 8, 12 and 16 Walsh Transform sectors with absolute Difference respectively. Figure9-12 shows the overall average cross over performance of individual sector mean of 4, 8, 12 and 16 Walsh sectors. The comparison bar chart of cross over points of overall average of precision and recall for 4, 8, 12 and 16 sectors of density distribution approach w.r.t. two different similarity measures namely Euclidean distance and Absolute difference is shown in the Figure13. It is observed that performance of 12 sectors is the best. The performance of absolute difference is quite close to Euclidean distance. The bar chart shown in the Figure14 compares the performance of individual sector mean of 4, 8, 12 and 16 Walsh sectors. It is observed that the sector 16 with this approach has the best outcome as 0.511 and 0.524 of cross over point with Euclidean distance and absolute difference similarity measures respectively.

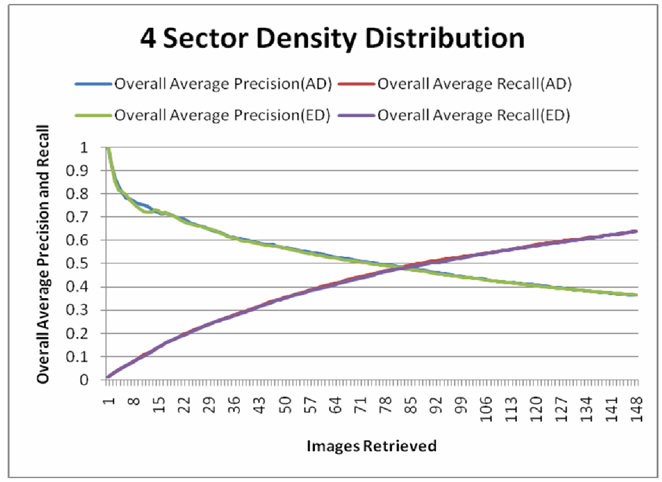


Figure 5: Overall Average Precision and Recall performance of sal cal density distribution in 4 Walsh Transform sectors with Absolute Difference(AD) and Euclidian Distance (ED) as similarity measures.

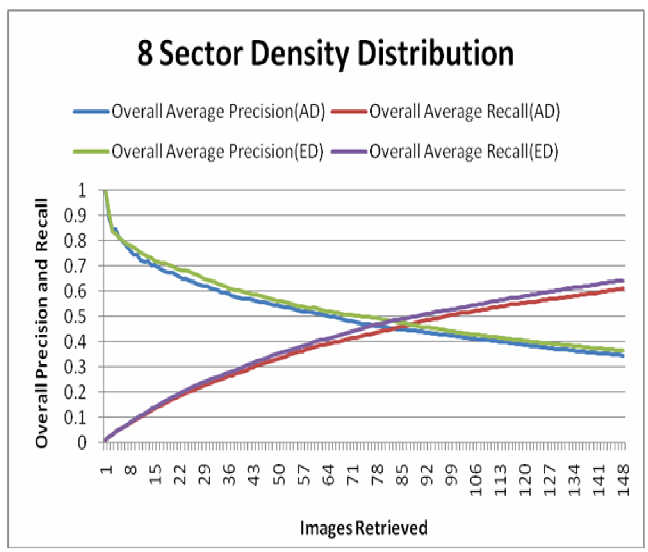


Figure 6: Overall Average Precision and Recall performance of sal cal density distribution in 8 Walsh Transform sectors with Absolute Difference(AD) and Euclidian Distance (ED) as similarity measures.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

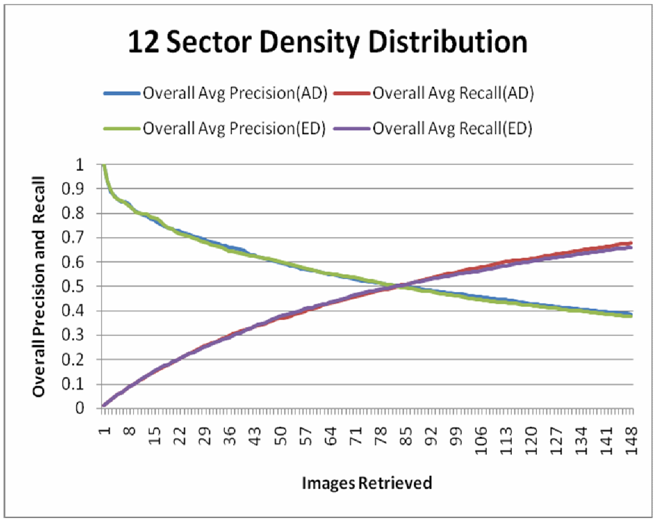


Figure 7: Overall Average Precision and Recall performance of sal cal density distribution in 12 Walsh Transform sectors with Absolute Difference(AD) and Euclidian Distance (ED) as similarity measures.

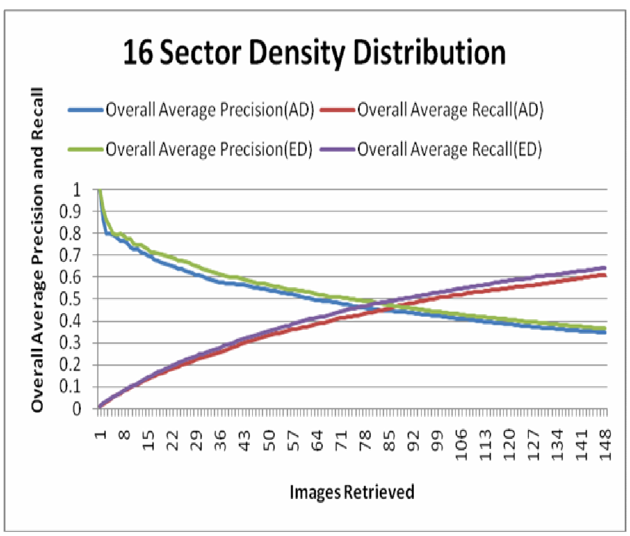


Figure 8: Overall Average Precision and Recall performance of sal cal density distribution in 16 Walsh Transform sectors with Absolute Difference(AD) and Euclidian Distance (ED) as similarity measures.

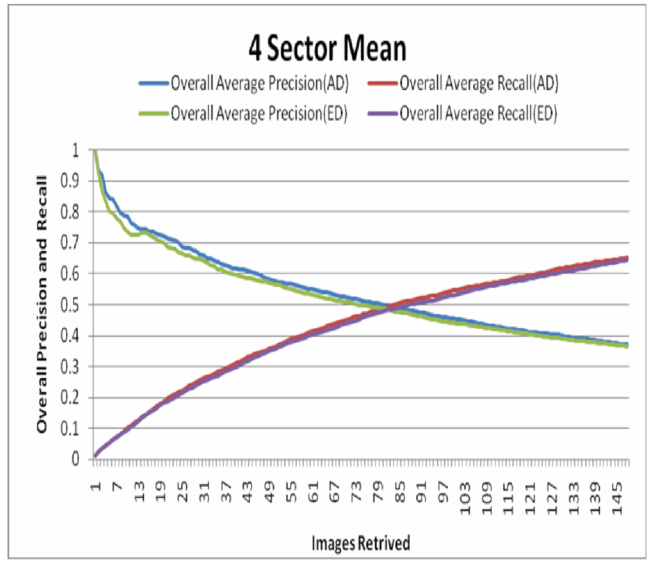


Figure 9: Overall Average Precision and Recall performance of Individual sector mean in 4 Walsh Transform sectors with Absolute Difference(AD) and Euclidian Distance (ED) as similarity measures.

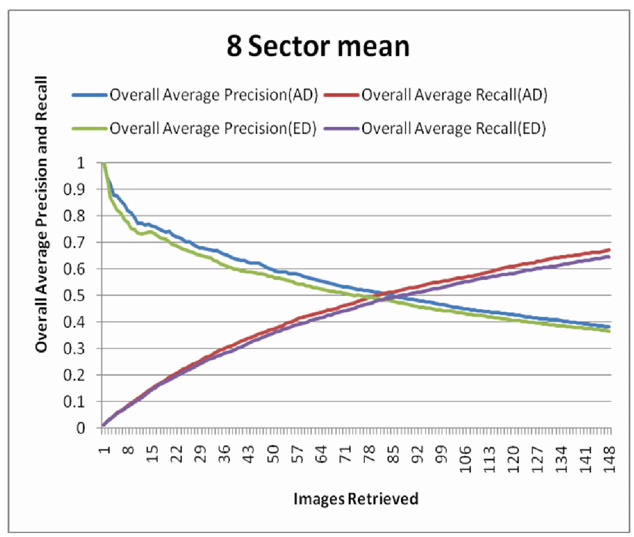


Figure 10: Overall Average Precision and Recall performance of Individual sector mean in 8 Walsh Transform sectors with Absolute Difference(AD) and Euclidian Distance (ED) as similarity measures.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

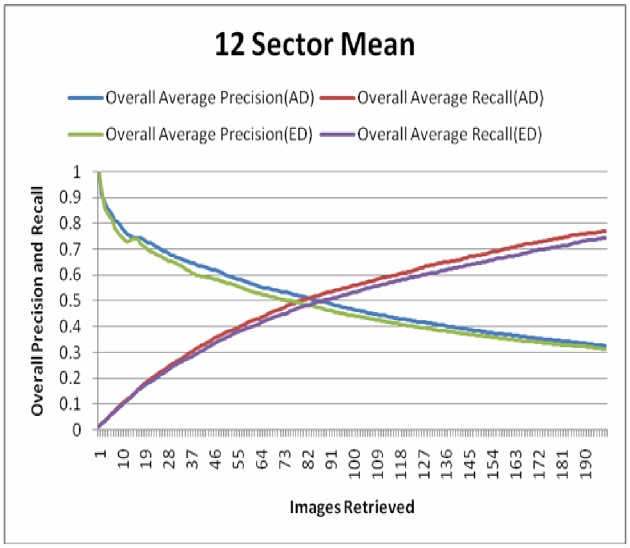


Figure 11: Overall Average Precision and Recall performance of Individual sector mean in 12 Walsh Transform sectors with Absolute Difference(AD) and Euclidian Distance (ED) as similarity measures.

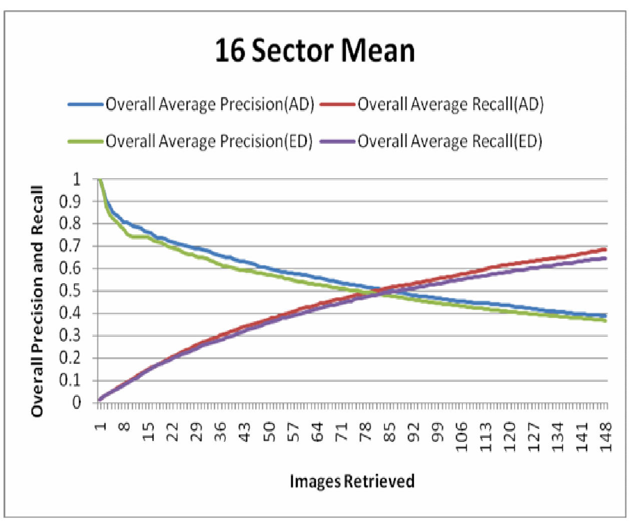


Figure 12: Overall Average Precision and Recall performance of Individual sector mean in 16 Walsh Transform sectors with Absolute Difference(AD) and Euclidian Distance (ED) as similarity measures.

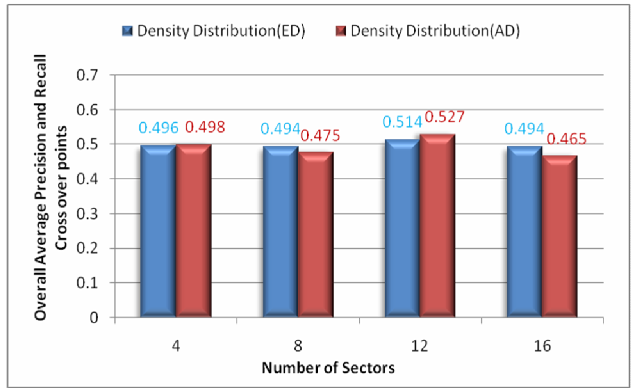


Figure 13: Comparison of Overall Precision and Recall cross over points based on sal cal density distribution in Walsh 4, 8, 12 and 16 sectors with Absolute Difference (AD) and Euclidean Distance (ED) as similarity measure.

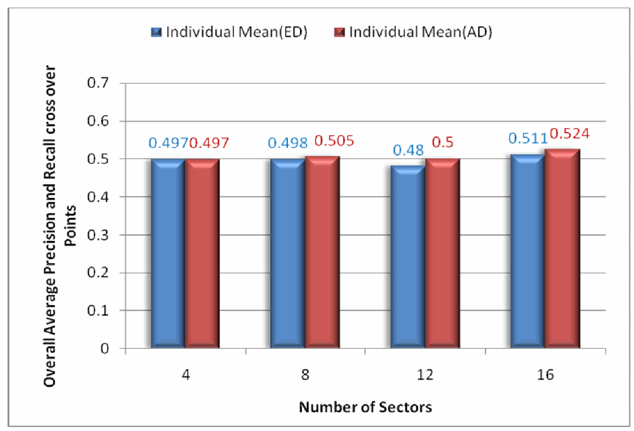


Figure 14: Comparison of Overall Precision and Recall cross over points based on individual sector mean in Walsh 4, 8, 12 and 16 sectors with Absolute Difference (AD) and Euclidean Distance (ED) as similarity measure.

.V. CONCLUSION

The Innovative idea of using complex Walsh transform 4, 8, 12 and 16 sectors of the images to generate the feature vectors for content based image retrieval is proposed. We have proposed two different approaches for feature vector generation with sum of absolute difference as similarity measuring parameter. First one is sal-cal density distribution in complex transform and second is individual sector mean. In both approaches we have used zero and highest sequency of cal and sal as extra components to augment the feature vector. This approach gives improvisation in the result of density distribution approach. It has increased the discretionary power of sector 4 which was quite absent in the case of non utilization of extra components of sequency [9].The cross over point performance of overall average of precision and recall for both approaches on all applicable sectors are compared. Using Walsh transform and sum of absolute difference as similarity measuring parameter reduces the computational complexity reducing the search time and calculation [8][9]. The density

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

distribution approach of feature vector augmentation with 12 sectors gives the best result of overall precision and recall cross over point of 0.514 and 0.527 for Euclidean distance and sum of absolute difference respectively. In case of individual sector mean approach the sector 16 gives the best result as 0.511 and 0.524 as cross over point of precision and recall with same similarity measures. Thus the best performance is obtained with density and sum of absolute difference as similarity measure which is computationally most economical.

REFERENCES

[1] Kato, T., "Database architecture for content based image retrieval in Image Storage and Retrieval Systems" (Jambardino A and Niblack W eds),*Proc SPIE 2185*, pp 112-123, 1992.

[2] John Berry and David A. Stoney "The history and development of fingerprinting," in *Advances in Fingerprint Technology*, Henry C. Lee and R. E. Gaensslen, Eds., pp. 1-40. CRC Press Florida, 2nd edition, 2001.

[3] Emma Newham, "The biometric report," SJB Services, 1995.

[4] H. B. Kekre, Dhirendra Mishra, "Digital Image Search & Retrieval using FFT Sectors" published in proceedings of National/Asia pacific conference on Information communication and technology(NCICT 10) 5TH & 6TH March 2010.SVKM'S NMIMS MUMBAI

[5] H.B.Kekre, Dhirendra Mishra, "Content Based Image Retrieval using Weighted Hamming Distance Image hash Value" published in the proceedings of international conference on contours of computing technology pp. 305-309 (Thinkquest2010) 13th & 14th March 2010.

[6] H.B.Kekre, Dhirendra Mishra,"Digital Image Search & Retrieval using FFT Sectors of Color Images" published in International Journal of Computer Science and Engineering (IJCSE) Vol. 02,No.02,2010,pp.368-372 ISSN 0975-3397 available online at http://www.enggjournals.com/ijcse/doc/IJCSE10-02-02-46.pdf

[7] H.B.Kekre, Dhirendra Mishra, "CBIR using upper six FFT Sectors of Color Images for feature vector generation" published in International Journal of Engineering and Technology(IJET) Vol. 02, No. 02, 2010, 49-54 ISSN 0975-4024 available online at http://www.enggjournals.com/ijet/doc/IJET10-02-02-06.pdf

[8] H.B.Kekre, Dhirendra Mishra, "Four walsh transform sectors feature vectors for image retrieval from image databases", published in international journal of computer sciecence and information technologies (IJCSIT) Vol. 1 (2) 2010, 33-37 ISSN 0975-9646 available online at http://www.ijcsit.com/docs/vol1issue2/ijcsit2010010201.pdf

[9] H.B.Kekre, Dhirendra Mishra, "Performance comparision of four, eight and twelve walsh transform sectors feature vectors for image retrieval from image databases", published in international journal of Engineering, science and technology(IJEST) Vol.2(5) 2010, 1370-1374 ISSN 0975-5462 available online at http://www.ijest.info/docs/IJEST10-02-05-62.pdf

[10] Arun Ross, Anil Jain, James Reisman, "A hybrid fingerprint matcher," *Int'l conference on Pattern Recognition (ICPR)*, Aug 2002.

[11] A. M. Bazen, G. T. B.Verwaaijen, S. H. Gerez, L. P. J. Veelenturf, and B. J. van der Zwaag, "A correlation-based fingerprint verification system," *Proceedings of the ProRISC2000 Workshop on Circuits, Systems and Signal Processing*, Veldhoven, Netherlands, Nov 2000.

[12] H.B.Kekre, Tanuja K. Sarode, Sudeep D. Thepade, "Image Retrieval using Color-Texture Features from DCT on VQ Codevectors obtained by Kekre's Fast Codebook Generation", ICGST International Journal on Graphics, Vision and Image Processing (GVIP), Available online at http://www.icgst.com/gvip

[13] H.B.Kekre, Sudeep D. Thepade, "Using YUV Color Space to Hoist the Performance of Block Truncation Coding for Image Retrieval", IEEE International Advanced Computing Conference 2009 (IACC'09), Thapar University, Patiala, INDIA, 6-7 March 2009.

[14] H.B.Kekre, Sudeep D. Thepade, "Image Retrieval using Augmented Block Truncation Coding Techniques", ACM International Conference on Advances in Computing, Communication and Control (ICAC3-2009), pp.: 384-390, 23-24 Jan 2009, Fr. Conceicao Rodrigous College of Engg., Mumbai. Available online at ACM portal.

[15] H.B.Kekre, Tanuja K. Sarode, Sudeep D. Thepade, "DCT Applied to Column mean and Row Mean Vectors of Image for Fingerprint Identification", International Conference on Computer Networks and Security, ICCNS-2008, 27-28 Sept 2008, Vishwakarma Institute of Technology, Pune.

[16] H.B.Kekre, Sudeep Thepade, Archana Athawale, Anant Shah, Prathmesh Velekar, Suraj Shirke, " Walsh transform over row mean column mean using image fragmentation and energy compaction for image retrieval", International journal of computer science and engineering (IJCSE),Vol.2.No.1,S2010,47-54.

[17] H.B.Kekre, Vinayak Bharadi, "Walsh Coefficients of the Horizontal & Vertical Pixel Distribution of Signature Template", In Proc. of Int. Conference ICIP-07, Bangalore University, Bangalore. 10-12 Aug 2007.

AUTHORS PROFILE

Dr. H. B. Kekre has received B.E. (Hons.) in Telecomm. Engg. from Jabalpur University in 1958, M.Tech (Industrial Electronics) from IIT Bombay in 1960, M.S.Engg. (Electrical Engg.) from University of Ottawa in 1965 and Ph.D.(System Identification) from IIT Bombay in 1970. He has worked Over 35 years as Faculty and H.O.D. Computer science and Engg. at IIT

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Bombay. From last 13 years working as a professor in Dept. of Computer Engg. at Thadomal Shahani Engg. College, Mumbai. He is currently senior Professor working with Mukesh Patel School of Technology Management and Engineering, SVKM's NMIMS University vile parle west Mumbai. He has guided 17 PhD.s 150 M.E./M.Tech Projects and several B.E./B.Tech Projects. His areas of interest are Digital signal processing, Image Processing and computer networking. He has more than 300 papers in National/International Conferences/Journals to his credit. Recently nine students working under his guidance have received the best paper awards. Recently two research scholars working under his guidance have completed PhD. Currently he is guiding 08 PhD. Students. He is life member of ISTE and Fellow of IETE.

Dhirendra S.Mishra has received his BE (Computer Engg) degree from University of Mumbai in 2002.Completed his M.E. (Computer Engg) from Thadomal shahani Engg. College, Mumbai, University of Mumbai. He is PhD Research Scholar and working as Assistant Professor in Computer Engineering department of Mukesh Patel School of Technology Management and Engineering, SVKM's NMIMS University, Mumbai, INDIA. He is life member of Indian Society of Technical education (ISTE), Member of International association of computer science and technology (IACSIT), Singapore, Member of International association of Engineers (IAENG). His areas of interests are Image Processing, Operating systems, Information Storage and Management

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

COMPARISON OF NEURAL NETWORK AND MULTIVARIATE DISCRIMINANT ANALYSIS IN SELECTING NEW COWPEA VARIETY

Adewole, Adetunji Philip *

Department of Computer Science, University of Agriculture, Abeokuta

philipwole@yahoo.com

Sofoluwe, A. B.

Department of Computer Science, University of Lagos, Akoka

Agwuegbo , Samuel Obi-Nnamdi

Department of Statistics, University of Agriculture, Abeokuta

E-mail: agwuegbo_son@yahoo.com

ABSTRACT

In this study, neural networks (NN) algorithm and multivariate discriminant (MDA) based model were developed to classify ten (10) varieties of cowpea which were widely planted in Kano. . In order to demonstrate the validity of our model, we use the case study to build a neural network model using Multilayer Feedforward Neural Network, and compare its classification performance against the Multivariate discriminant analysis. Two groups of data (Spray and Nospray) were used. Twenty kernels were used as training data set and test data to classify cowpea seed varieties. The neural network classified the new cowpea seed varieties based on the information it is trained with. At the end both methods were compared for their strength and weakness. It is noted that NN performed better than MDA, so that NN could be considered as a support tool in the process of selection of new cowpea varieties.

KEYWORDS: Cowpea, Multivariate Discriminant Analysis (MDA), Neural Network (NN), Perceptron.

1.0 Introduction

The history of neural networks begins with the earliest model of the biological neuron given by [5]. This model describes a neuron as a linear threshold computing unit with multiple inputs and a single output of either 0, if the nerve cell remains inactive, or 1, if the cell fires. A neuron fires if the sum of the inputs exceeds a specified threshold. In functional form, this gives $f(x) = 1$ for x greater than some threshold, and $f(x) = 0$ otherwise (this is commonly known as the indicator function). In theory, such a "system" of neurons presents a possible model for biological neural networks such as the human nervous system.

The [5] model was utilized in the development of the first artificial neural network by [12] in 1959. This network was based on a unit called the *perceptron*, which produces an output scaled as 1 or -1 depending upon the weighted, linear combination of inputs. Variations on the perceptron-based artificial neural network were further explored during the 1960s by [12] and by [15], among others.

In 1969 [6] demonstrated that the perceptron was incapable of representing simple functions which were linearly inseparable. This includes the case of the "exclusive or" (XOR). Because of this fundamental flaw (and Rosenblatt's untimely death) the study of neural networks fell into something of a decline during the 1970s. However, this limitation was overcome in the early 1980s. According to [11] : The post-perceptron era began with the realization that adding (hidden) layers to the network could yield significant computational versatility. This yielded a considerable revival of interest in ANNs (especially multilayered feedforward structures), which continues to this day.

Presently, much research on neural networks is taking place within two areas: (a) the aforementioned multilayered feed-forward networks, also known as multilayer perceptrons, and (b) symmetric recurrent networks, also known as attractor neural networks or Hopfield nets. The former model is used for classification problems, while the latter is used for developing associative memory systems. The investigation into neural network structures and performance has taken on a substantially pragmatic feel in recent years. There is greater interest in using neural networks as problem-solving algorithms than in developing them as accurate representations of the human nervous system. ANNs have been implemented to solve a variety of problems involving pattern classification and pattern recognition.

2.0 Neural Networks and Standard Statistical Techniques

Similarities between ANNs and statistical methods definitely exist. Indeed, neural networks have been categorized as a form of nonlinear regression. It has also been observed that multiple linear regression, a standard statistical tool, can be expressed in terms of a simple ANN node. For example, given the linear equation $y = b_0 + b_1x_1 + ... + b_nx_n$, the x_i can be taken as the inputs to a node, the b_i taken as the corresponding weights, and b_0 taken as the activation function. There are at least two key differences between ANNs and statistical methods. Often remarked upon as a major drawback of ANNs is the fact that their internal functional structure remains unknown once they have been trained. In effect, a neural network remains a "black box" that may produce useful results, but cannot be precisely understood. Statistical procedures do not exhibit this sort of opaque design. The construction of a neural network is also something of an ad-hoc process whereas there are commonly formalized guidelines for fitting the best model in statistics. The performance of ANNs has been extensively compared to that of various statistical methods within the areas of prediction and

.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

classification. In particular, a fair amount of literature has been generated on the use of ANNs in time series forecasting. In an examination of two time series without noise, [4] concluded that basic neural networks substantially outperform conventional statistical methods. [7] found that a neural network and the Box-Jenkins forecasting system performed about the same for the analysis of 75 different time series. Interestingly, the memory of a time series has been demonstrated to influence relative performance of ANNs and the Box-Jenkins Model. The Box-Jenkins model slightly outperforms ANNs for time series with long memory, while the reverse tends to be true for time series with short memory. Stern has also concluded that for time series analysis "NNs work best in problems with little or no stochastic component". Neural network and statistical approaches to pattern classification have been compared by a number of researchers. For the most part, reviews seem to be mixed. For instance, [3] concluded that neural networks show little promise as real-world classifiers, while a case study examined by Yoon points to the superiority of ANNs over classical discriminant analysis.

In a comprehensive study of classification techniques, [6] rated the performance of a large selection of neural network, statistical, and machine learning algorithms on a variety of data sets. In the analysis of their results, they presented the top five algorithms for twenty-two different data sets based on error rates. Though not conclusive, the study by Mitchie would seem to suggest that neural networks aren't necessarily replacements for, or even preferable alternatives to standard statistical classification techniques.

2.1 Multivariate Discriminant Analysis

The term multivariate discriminant analysis refers to several different types of analyses. Classificatory discriminant analysis is used to classify observations into two or more known groups on the basis of one or more quantitative variables. Classification can be done by either a parametric method or a nonparametric method. A parametric method is appropriate only for approximating normal within-class distributions. The method generates either a linear discriminant function or a quadratic discriminant function. When the distribution within each group is not assumed to have any specific distribution different from the multivariate normal distribution, nonparametric methods can be used to derive classification criteria. These methods include the kernel method and nearest-neighbor methods. The kernel method uses uniform, normal, bi-weight, or tri-weight kernels in estimating the group-specific density at each observation. The within-group covariance matrices or the pooled covariance matrix can be used to scale the data. The performance of a discriminant function can be evaluated by estimating error rates (probabilities of misclassification). Error count estimates and posterior probability error rate estimates can be evaluated. In multivariate statistical applications, the data collected are largely from distributions different from the normal distribution. Various forms of nonnormality can arise, such as qualitative variables or variables with underlying continuous but nonnormal distributions. If the multivariate normality assumption is violated, the use of parametric discriminant analysis might not be appropriate. When a parametric classification criterion (linear or quadratic discriminant function) is derived from a nonnormal population, the resulting error rate estimates might be biased.

2.1.1 Discriminant Function

A simple linear discriminant function transforms an original set of measurements on a sample into a single discriminant score. The score or transform variable represents the sample's position along line defined by the linear discriminant function. We can therefore think of the discriminant function as away of collapsing a multivariate problem down into a problem which involves only one

variable. One method that can be used to find discriminant function is regression; the dependent variable consists of the differences between the multivariate means of the group.

In matrix notation, we must solve an equation of the form

$[S_p^2]*[\lambda] = [D]$………………………………… (1)

Where $\mathbf{S_p^2}$ is the m*m matrix of pooled variance and covariances of the m- variable, $\boldsymbol{\lambda}$ is the coefficient of the D-equation.

$[\lambda] = [S_p^2]^{-1}*[D]$………………………………… (2)

To compute the discriminant function, we must determine the various entries in the matrix equation. The mean differences are found simply by

$D_j=\hat{A}_j-B_j=\sum_{i=1}^{na} Aij/n_a-\sum_{i=1}^{na} Bij/n_b$ …………………… (3)

In expanded form

$$\begin{pmatrix} D_1 \\ D_2 \\ - \\ - \\ - \\ D_m \end{pmatrix} = \begin{pmatrix} \hat{A}_1 \\ \hat{A}_2 \\ - \\ - \\ - \\ A_m \end{pmatrix} - \begin{pmatrix} B_1 \\ B_2 \\ - \\ - \\ - \\ B_m \end{pmatrix}$$

We must construct the matrix of sum of square and cross product of all variables.

Sum of product of Matrix

$A = \sum_{i=1}^{na}(AijAik) - (\sum_{i=1}^{na} Aij \sum_{i=1}^{na} Aik)/na$……. (4)

Sum of product of Matrix

$B = \sum_{i=1}^{nb}(BijBik) - (\sum_{i=1}^{nb} Bij \sum_{i=1}^{nb} Bik)/nb$…… (5)

Then the matrix of pooled variance can be found as

$(S_p^2) = (SPA + SPB)/(na + nb - 2)$………….. (6)

It could be observed that this equation for pooled variance is exactly the same as that used in the T^2 test of the equality of multivariate means. Although the amounts of mathematical manipulation that must be performed to calculate the coefficients of a discriminant function appear large.

.

 http://sites.google.com/site/ijcsis/
ISSN 1947-5500

The set of coefficients that can be found are entries in the discriminant function equation of the form:

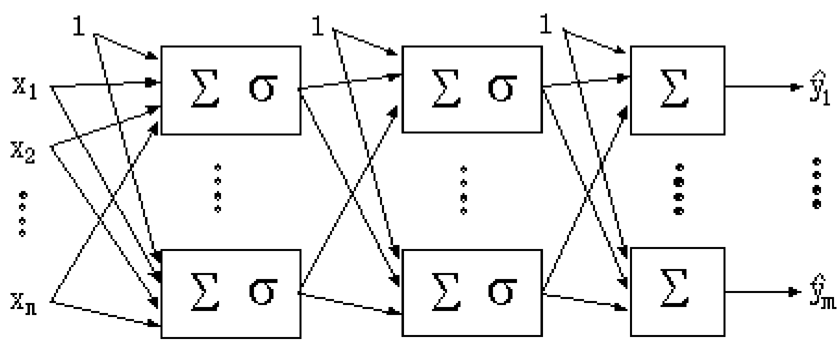
$R_0 = \lambda_1\psi_1 + \lambda_2\psi_2 + \lambda_3\psi_3 + \ldots\ldots + \lambda_m\psi \ldots\ldots\ldots\ldots\ldots\ldots\ldots$ (7)

The discriminant index, R_0, is the point along the discriminant function line which is exactly halfway between the center of group A and the center of group B. Then we substitute the multivariate mean of group A into the equation to obtain R_A (that is, we set $\psi_j = A_j$) and the mean of group B.

2.2 Multilayer Feedforward Neural Network

Feedforward neural networks (FF networks) are the most popular and most widely used models in many practical applications. Feed-forward ANNs allow signals to travel one way only; from input to output. There is no feedback (loops) i.e. the output of any layer does not affect that same layer. Feed-forward ANNs tend to be straight forward networks that associate inputs with outputs. They are extensively used in pattern recognition. This type of organisation is also referred to as bottom-up or top-down. They are known by many different names, such as "multi-layer perceptrons. The manner in which the neurons of a neural network are structured is intimately linked with the learning algorithm used to train the network. We may therefore speak of learning algorithm used in the design of neural networks as being structured. The classification of learning algorithms can be considered based on two fundamental different classes (Single layer feed-forward, Multi-layer feed-forward) of neural network architecture.

The diagram bellow describes the structure of the multilayer feed-forward neural network.



Input Layer Hidden Layer Output Layer

Fig1

a. The activity of the input units represents the raw information that is fed into the network.
b. The activity of each hidden unit is determined by the activities of the input units and the weights on the connections between the input and the hidden units.
c. The behavior of the output units depends on the activity of the hidden units and the weights between the hidden and output units. This simple type of network is interesting because the hidden units are free to construct their own representations of the input. The weights between the input and hidden units determine when each hidden unit is active, and so by modifying these weights, a hidden unit can choose what it represents.

3.0 Implementation

A package was developed and implemented using java and S-plus as the computing environment and run a on Pentium IV 1.80GHz of processor, 512MB of RAM and 40GB of local disk. These packages can work on any windows based operating system but for best performance, window XP professional is recommended.

The implementation could be started by lunching into the interface. It is followed by the menu editor which gives user access to the whole application of neural network. The two classes (No Spray, Spray) of the data were first tested for the similarity/dissimilarity and correlation test was performed to generate the training weights from the given data using S-Plus, and the results were shown below:

Table: Results of the analysis.

Variable	RankSumSquare(RSS)	Weight	Std.Error	Intercept
Good Seed	84900.2	0.52	0.000000+00	spray.gs
Bad Seed	84900.2	0.45	0.000000+00	spray.bs
Germinating Seed	1824.32	0.40	0.000000+00	spray.sw
Seed Weight	7339.87	0.25	0.000000+00	spray.sg
Days to flowering	12295.91	-0.31	0.000000+00	spray.df
Days to maturity	1828.34	0.27	0.000000+00	spray.dm
Plant stand/hectare	20272.38	0.24	0.000000+00	spray.sh

Degree of freedom: 30 totals; 28 residual

Residual standard error (on weighted scale): 55.06496

The model for the correlation and the neural network are as follow:

Y = Cowpea yield

$$Y = \sum_1^7 bixi$$

$$Y_{ij} = \sum_1^7 WijXi + \text{Error}$$

Where $W_{ij} = (X''X)^{-1}X''Y$

The figure shows the application as it appears when it initially starts up. Initially, the network activation weights are all zero. The network has learned no patterns at this point.

.

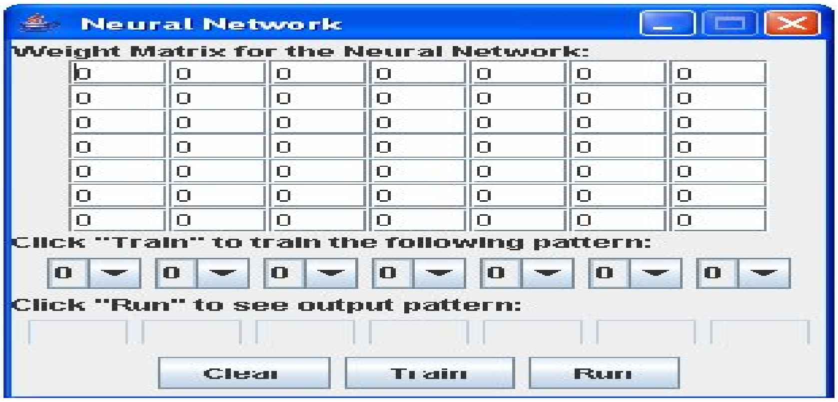


Fig2: neural network before training

To train the network to recognize the pattern 1000000, enter 1000000 under the "Input pattern to run or train". Click the "Train" button. Notice the weight matrix adjust to absorb the new knowledge.

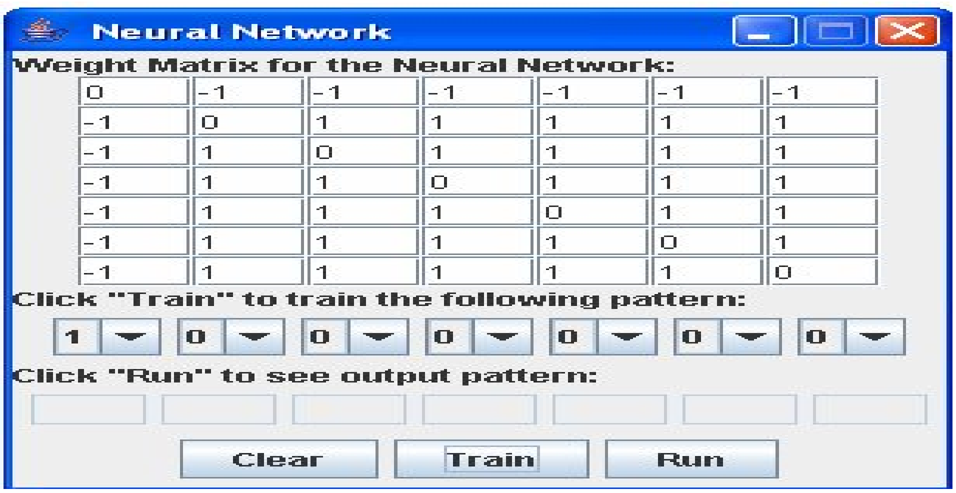


Fig3: neural network after the first training

Now the network will be tested. Enter the pattern 1000000 into the "Input pattern to run or train" (it should still be there from your training). The output will be "1000000". This is an auto associative network, therefore it echoes the input if it recognizes it.

Now try something that does not match the training pattern exactly. Enter the pattern "01000000" and click "Run". The output will now be "0111000". The neural network did not recognize "01000000", but the closest thing it knew was "0111000". Now test the side effect mentioned previously. Enter "0111111", which is the binary inverse of what the network was trained with ("0111111"). The networks always get trained for the binary inverse too. So if you enter "00000001", the network will recognize it.

Then , the final test. Enter "0000000", which is totally off base and not close to anything the neural network knows. The neural network responds with "0111000", it did try to correct, it has idea what you mean. You can play with the network more. It can be

.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

taught more than one pattern. As you train new patterns it builds upon the matrix already in memory. Pressing “Clear” clears out the memory.

4.0 Discussion of Result

Table 2:

Ten (10) varieties of Cowpea were used for classification and the results are shown bellow:

Variety	MDA	Neural Network	Test
IT845-224-4	A	A	A
IT86D-716	R	A	A
IT90K-277-2	A	A	A
Tvu-13743	A	A	A
Tvu-1890	A	A	A
Tvu-14476	A	A	A
IT86D-715	A	A	A
IT86D-719	A	A	A
Tvu13731	A	A	A
TvNu72	A	A	A

A = Accept

R = Reject

For the NN, the results may be considered very promising. In ten (10) varieties that were tested, all the varieties were accepted in the test, one (IT86D-716) is rejected by the multivariate discriminant but accepted by the neural network.

5.0 Conclusion

To explore new ways to help the process of selection of new cowpea varieties, neural networks (NN) algorithm and multivariate discriminant (MDA) based model were developed to classify ten (10) varieties of cowpea which were widely planted in Kano and, the methods (NN and MDA) used showed that Neural Networks could be considered as a promising technique to develop support tools for the process of selection of new cowpea varieties. For the NN, the results may be considered very promising. In ten (10) varieties that were tested, all the varieties were accepted in the test, one (IT86D-716) is rejected by the multivariate discriminant but accepted by the neural network.

REFERENCES

[1] Alan S. Lapedes, Robert M. Farber: How Neural Nets Work. NIPS 1987: 442-456

[2] Elisa R. and Marco B. (2007). Multivariate Statical Tools for the Evaluation of Proteomic 2D-maps. Protomics 4:53-66.

[3] Jyhshyan Lan, Michael Y. Hu, B. Eddy Patuwo, G. Peter Zhang: An investigation of neural network classifiers with unequal misclassification costs and group sizes. Decision Support Systems 48(4): 582-591 (2010)

[4] **Lapedes**, A. & **Farber**, R. (1987): Nonlinear signal processing using neural network. International **Journal** of Forecasting. v14. 323-337.

[5] McCulloch W.S. and Pitts W. (1943), "A Logical Calculus of the Ideas Immanent in Nervous Activity," Bulletin of Mathematical Biophysics, 5, 115-133.

[6] Minsky M. and Papert S. (1969): *Perceptrons: An Introduction to Computational Geometry*. MIT Press, Cambridge, Mass.

[7] Patil, P. N. and Ratnaparkhi, M. V. (2000). A nonparametric test for size biasedness. Jour. of Indian Statist. Assoc., 38, 369-382.

[8] Patil, P. N. and Speckman, P. (2004). Constrained kernel regression. Jour. of Indian Statist. Assoc., 42, 87-98.

[9] Potts W.J.E. (2000), Neural Network Modeling Course Notes, Cary: SAS Institute, Inc.

[10] Quian N and Sejnowski T.J. (1988). "Predicting the Secondary Structure of Globular Proteins Using Neural Network Models," Journal of Molecular Biology, 202, 865-884.

[11] Robert J. Schalkoff: (1987): Analysis of the weak solution approach to image motion estimation. Pattern Recognition 20 (2): 189-197.

[12] Rosenblatt, F. (1958). "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain," *Psychol. Rev.,* Vol. 65, p. 386.

[13] Rosenblatt, F. (1960). "Perceptron Simulation Experiments," *Proc. IRE,* Vol. 48, pp. 301–309.

[14] Trumbo B.E. Norton J.A. Freerks L. (1999), "Using CIS/ED to Make a Reading List on Neural Nets," STATs Magazine, winter, 1999.

[15] Widrow 1. B. and. Hoff M. E, Jr. (1960). "Adaptive Switching Circuits," IRE WESCON Conv. Rec., Part 4, pp. 96–104.

[16] Zhang, G., B. Eddy Patuwo, et al. (1998). "Forecasting with artificial neural networks: The state of the art." International Journal of Forecasting 14(1): 35-62.

.

IJCSIS REVIEWERS' LIST

CALL FOR PAPERS
International Journal of Computer Science and Information Security
IJCSIS 2010

ISSN: 1947-5500
http://sites.google.com/site/ijcsis/

International Journal Computer Science and Information Security, now at its sixth edition, is the premier scholarly venue in the areas of computer science and security issues. IJCSIS 2010 will provide a high profile, leading edge platform for researchers and engineers alike to publish state-of-the-art research in the respective fields of information technology and communication security. The journal will feature a diverse mixture of publication articles including core and applied computer science related topics.

Authors are solicited to contribute to the special issue by submitting articles that illustrate research results, projects, surveying works and industrial experiences that describe significant advances in the following areas, but are not limited to. Submissions may span a broad range of topics, e.g.:

Track A: Security

Access control, Anonymity, Audit and audit reduction & Authentication and authorization, Applied cryptography, Cryptanalysis, Digital Signatures, Biometric security, Boundary control devices, Certification and accreditation, Cross-layer design for security, Security & Network Management, Data and system integrity, Database security, Defensive information warfare, Denial of service protection, Intrusion Detection, Anti-malware, Distributed systems security, Electronic commerce, E-mail security, Spam, Phishing, E-mail fraud, Virus, worms, Trojan Protection, Grid security, Information hiding and watermarking & Information survivability, Insider threat protection, Integrity
Intellectual property protection, Internet/Intranet Security, Key management and key recovery, Language-based security, Mobile and wireless security, Mobile, Ad Hoc and Sensor Network Security, Monitoring and surveillance, Multimedia security ,Operating system security, Peer-to-peer security, Performance Evaluations of Protocols & Security Application, Privacy and data protection, Product evaluation criteria and compliance, Risk evaluation and security certification, Risk/vulnerability assessment, Security & Network Management, Security Models & protocols, Security threats & countermeasures (DDoS, MiM, Session Hijacking, Replay attack etc,), Trusted computing, Ubiquitous Computing Security, Virtualization security, VoIP security, Web 2.0 security, Submission Procedures, Active Defense Systems, Adaptive Defense Systems, Benchmark, Analysis and Evaluation of Security Systems, Distributed Access Control and Trust Management, Distributed Attack Systems and Mechanisms, Distributed Intrusion Detection/Prevention Systems, Denial-of-Service Attacks and Countermeasures, High Performance Security Systems, Identity Management and Authentication, Implementation, Deployment and Management of Security Systems, Intelligent Defense Systems, Internet and Network Forensics, Large-scale Attacks and Defense, RFID Security and Privacy, Security Architectures in Distributed Network Systems, Security for Critical Infrastructures, Security for P2P systems and Grid Systems, Security in E-Commerce, Security and Privacy in Wireless Networks, Secure Mobile Agents and Mobile Code, Security Protocols, Security Simulation and Tools, Security Theory and Tools, Standards and Assurance Methods, Trusted Computing, Viruses, Worms, and Other Malicious Code, World Wide Web Security, Novel and emerging secure architecture, Study of attack strategies, attack modeling, Case studies and analysis of actual attacks, Continuity of Operations during an attack, Key management, Trust management, Intrusion detection techniques, Intrusion response, alarm management, and correlation analysis, Study of tradeoffs between security and system performance, Intrusion tolerance systems, Secure protocols, Security in wireless networks (e.g. mesh networks, sensor networks, etc.), Cryptography and Secure Communications, Computer Forensics, Recovery and Healing, Security Visualization, Formal Methods in Security, Principles for Designing a Secure Computing System, Autonomic Security, Internet Security, Security in Health Care Systems, Security Solutions Using Reconfigurable Computing, Adaptive and Intelligent Defense Systems, Authentication and Access control, Denial of service attacks and countermeasures, Identity, Route and

Location Anonymity schemes, Intrusion detection and prevention techniques, Cryptography, encryption algorithms and Key management schemes, Secure routing schemes, Secure neighbor discovery and localization, Trust establishment and maintenance, Confidentiality and data integrity, Security architectures, deployments and solutions, Emerging threats to cloud-based services, Security model for new services, Cloud-aware web service security, Information hiding in Cloud Computing, Securing distributed data storage in cloud, Security, privacy and trust in mobile computing systems and applications, **Middleware security & Security features:** middleware software is an asset on
its own and has to be protected, interaction between security-specific and other middleware features, e.g., context-awareness, **Middleware-level security monitoring and measurement:** metrics and mechanisms for quantification and evaluation of security enforced by the middleware, **Security co-design:** trade-off and co-design between application-based and middleware-based security, **Policy-based management:** innovative support for policy-based definition and enforcement of security concerns, **Identification and authentication mechanisms:** Means to capture application specific constraints in defining and enforcing access control rules, **Middleware-oriented security patterns:** identification of patterns for sound, reusable security, **Security in aspect-based middleware:** mechanisms for isolating and enforcing security aspects, **Security in agent-based platforms:** protection for mobile code and platforms, Smart Devices: Biometrics, National ID cards, Embedded Systems Security and TPMs, RFID Systems Security, Smart Card Security, Pervasive Systems: Digital Rights Management (DRM) in pervasive environments, Intrusion Detection and Information Filtering, Localization Systems Security (Tracking of People and Goods), Mobile Commerce Security, Privacy Enhancing Technologies, Security Protocols (for Identification and Authentication, Confidentiality and Privacy, and Integrity), Ubiquitous Networks: Ad Hoc Networks Security, Delay-Tolerant Network Security, Domestic Network Security, Peer-to-Peer Networks Security, Security Issues in Mobile and Ubiquitous Networks, Security of GSM/GPRS/UMTS Systems, Sensor Networks Security, Vehicular Network Security, Wireless Communication Security: Bluetooth, NFC, WiFi, WiMAX, WiMedia, others

This Track will emphasize the design, implementation, management and applications of computer communications, networks and services. Topics of mostly theoretical nature are also welcome, provided there is clear practical potential in applying the results of such work.

Track B: Computer Science

Broadband wireless technologies: LTE, WiMAX, WiRAN, HSDPA, HSUPA, Resource allocation and interference management, Quality of service and scheduling methods, Capacity planning and dimensioning, Cross-layer design and Physical layer based issue, Interworking architecture and interoperability, Relay assisted and cooperative communications, Location and provisioning and mobility management, Call admission and flow/congestion control, Performance optimization, Channel capacity modeling and analysis, Middleware Issues: Event-based, publish/subscribe, and message-oriented middleware, Reconfigurable, adaptable, and reflective middleware approaches, Middleware solutions for reliability, fault tolerance, and quality-of-service, Scalability of middleware, Context-aware middleware, Autonomic and self-managing middleware, Evaluation techniques for middleware solutions, Formal methods and tools for designing, verifying, and evaluating, middleware, Software engineering techniques for middleware, Service oriented middleware, Agent-based middleware, Security middleware, Network Applications: Network-based automation, Cloud applications, Ubiquitous and pervasive applications, Collaborative applications, RFID and sensor network applications, Mobile applications, Smart home applications, Infrastructure monitoring and control applications, Remote health monitoring, GPS and location-based applications, Networked vehicles applications, Alert applications, Embeded Computer System, Advanced Control Systems, and Intelligent Control : Advanced control and measurement, computer and microprocessor-based control, signal processing, estimation and identification techniques, application specific IC's, nonlinear and adaptive control, optimal and robot control, intelligent control, evolutionary computing, and intelligent systems, instrumentation subject to critical conditions, automotive, marine and aero-space control and all other control applications, Intelligent Control System, Wiring/Wireless Sensor, Signal Control System. Sensors, Actuators and Systems Integration : Intelligent sensors and actuators, multisensor fusion, sensor array and multi-channel processing, micro/nano technology, microsensors and microactuators, instrumentation electronics, MEMS and system integration, wireless sensor, Network Sensor, Hybrid

Sensor, Distributed Sensor Networks. Signal and Image Processing : Digital signal processing theory, methods, DSP implementation, speech processing, image and multidimensional signal processing, Image analysis and processing, Image and Multimedia applications, Real-time multimedia signal processing, Computer vision, Emerging signal processing areas, Remote Sensing, Signal processing in education. Industrial Informatics: Industrial applications of neural networks, fuzzy algorithms, Neuro-Fuzzy application, bioInformatics, real-time computer control, real-time information systems, human-machine interfaces, CAD/CAM/CAT/CIM, virtual reality, industrial communications, flexible manufacturing systems, industrial automated process, Data Storage Management, Harddisk control, Supply Chain Management, Logistics applications, Power plant automation, Drives automation. Information Technology, Management of Information System : Management information systems, Information Management, Nursing information management, Information System, Information Technology and their application, Data retrieval, Data Base Management, Decision analysis methods, Information processing, Operations research, E-Business, E-Commerce, E-Government, Computer Business, Security and risk management, Medical imaging, Biotechnology, Bio-Medicine, Computer-based information systems in health care, Changing Access to Patient Information, Healthcare Management Information Technology. Communication/Computer Network, Transportation Application : On-board diagnostics, Active safety systems, Communication systems, Wireless technology, Communication application, Navigation and Guidance, Vision-based applications, Speech interface, Sensor fusion, Networking theory and technologies, Transportation information, Autonomous vehicle, Vehicle application of affective computing, Advance Computing technology and their application : Broadband and intelligent networks, Data Mining, Data fusion, Computational intelligence, Information and data security, Information indexing and retrieval, Information processing, Information systems and applications, Internet applications and performances, Knowledge based systems, Knowledge management, Software Engineering, Decision making, Mobile networks and services, Network management and services, Neural Network, Fuzzy logics, Neuro-Fuzzy, Expert approaches, Innovation Technology and Management : Innovation and product development, Emerging advances in business and its applications, Creativity in Internet management and retailing, B2B and B2C management, Electronic transceiver device for Retail Marketing Industries, Facilities planning and management, Innovative pervasive computing applications, Programming paradigms for pervasive systems, Software evolution and maintenance in pervasive systems, Middleware services and agent technologies, Adaptive, autonomic and context-aware computing, Mobile/Wireless computing systems and services in pervasive computing, Energy-efficient and green pervasive computing, Communication architectures for pervasive computing, Ad hoc networks for pervasive communications, Pervasive opportunistic communications and applications, Enabling technologies for pervasive systems (e.g., wireless BAN, PAN), Positioning and tracking technologies, Sensors and RFID in pervasive systems, Multimodal sensing and context for pervasive applications, Pervasive sensing, perception and semantic interpretation, Smart devices and intelligent environments, Trust, security and privacy issues in pervasive systems, User interfaces and interaction models, Virtual immersive communications, Wearable computers, Standards and interfaces for pervasive computing environments, Social and economic models for pervasive systems, Active and Programmable Networks, Ad Hoc & Sensor Network, Congestion and/or Flow Control, Content Distribution, Grid Networking, High-speed Network Architectures, Internet Services and Applications, Optical Networks, Mobile and Wireless Networks, Network Modeling and Simulation, Multicast, Multimedia Communications, Network Control and Management, Network Protocols, Network Performance, Network Measurement, Peer to Peer and Overlay Networks, Quality of Service and Quality of Experience, Ubiquitous Networks, Crosscutting Themes – Internet Technologies, Infrastructure, Services and Applications; Open Source Tools, Open Models and Architectures; Security, Privacy and Trust; Navigation Systems, Location Based Services; Social Networks and Online Communities; ICT Convergence, Digital Economy and Digital Divide, Neural Networks, Pattern Recognition, Computer Vision, Advanced Computing Architectures and New Programming Models, Visualization and Virtual Reality as Applied to Computational Science, Computer Architecture and Embedded Systems, Technology in Education, Theoretical Computer Science, Computing Ethics, Computing Practices & Applications

Authors are invited to submit papers through e-mail ijcsiseditor@gmail.com. Submissions must be original and should not have been published previously or be under consideration for publication while being evaluated by IJCSIS. Before submission authors should carefully read over the journal's Author Guidelines, which are located at http://sites.google.com/site/ijcsis/authors-notes .

ISSN 1947 5500